

COMPOSITIONS AND METHODS RELATING TO OSTEOARTHRITIS

Field of the Invention

The invention relates to the profiling of differential gene expression in specific human tissue types through the construction and use of cDNA libraries and microarrays.

5 Background of the Invention

Osteoarthritis (OA) is a chronic disease in which the articular cartilage that lies on the ends of bones that forms the articulating surface of the joints gradually degenerates over time. There are many factors that are believed to predispose a patient to osteoarthritis including genetic susceptibility, obesity, accidental or athletic trauma, surgery, drugs and heavy physical demands. Osteoarthritis is initiated by damage to the cartilage of joints. The two most common injuries to joints are sports-related injuries and long term “repetitive use” joint injuries. Joints most commonly affected by osteoarthritis are the knees, hips and hands. In most cases, due to the essential weight-bearing function of the knees and hips, osteoarthritis in these joints causes much more disability than osteoarthritis of the hands. As cartilage degeneration progresses, secondary changes occur in other tissues in and around joints including bone, muscle, ligaments, menisci and synovium. The net effect of the primary failure of cartilage tissue and secondary damage to other tissues is that the patient experiences pain, swelling, weakness and loss of functional ability in the afflicted joint(s). These symptoms frequently progress to the point that they have a significant impact in terms of lost productivity and or quality of life consequences for the patient.

Articular cartilage is predominantly composed of chondrocytes, type II collagen, proteoglycans and water. Articular cartilage has no blood or nerve supply and chondrocytes are the only type of cell in this tissue. Chondrocytes are responsible for manufacturing the type II collagen and proteoglycans that form the cartilage matrix. This matrix in turn has physical-chemical properties that allow for saturation of the matrix with water. The net effect

of this structural-functional relationship is that articular cartilage has exceptional wear characteristics and allows for almost frictionless movement between the articulating cartilage surfaces. In the absence of osteoarthritis, articular cartilage often provides a lifetime of pain-free weight bearing and unrestricted joint motion even under demanding physical conditions.

5 During fetal development, articular cartilage is initially derived from the interzone of mesenchymal condensations. The mesenchymal cells cluster together and synthesize matrix proteins. The tissue is recognized as cartilage when the accumulation of matrix separates the cells, which are spherical in shape and are now called chondrocytes. During cartilage formation and growth, chondrocytes proliferate rapidly and synthesize large volumes of
10 matrix. Prior to skeletal maturity, chondrocytes are at their highest level of metabolic activity. As skeletal maturation is reached, the rate of chondrocyte metabolic activity and cell division declines. After completion of skeletal growth, most chondrocytes do not divide but do continue to synthesize matrix proteins such as collagens, proteoglycans and other noncollagenous proteins (1, 2).

15 Like all living tissues, articular cartilage is continually undergoing a process of renewal in which "old" cells and matrix components are being removed (catabolic activity) and "new" cells and molecules are being produced (anabolic activity). Relative to most tissues, the rate of anabolic/catabolic turnover in articular cartilage is low. Long-term maintenance of the structural integrity of mature cartilage relies on the proper balance
20 between matrix synthesis and degradation. Chondrocytes maintain matrix equilibrium by responding to chemical and mechanical stimuli from their environment. Appropriate and effective chondrocyte responses to these stimuli are essential for cartilage homeostasis. Disruption of homeostasis through either inadequate anabolic activity or excessive catabolic activity can result in cartilage degradation and osteoarthritis (3). Most tissues that are
25 damaged and have increased catabolic activity are able to mount an increased anabolic response that allows for tissue healing. Unfortunately, chondrocytes have very limited ability to up-regulate their anabolic activity and increase the synthesis of proteoglycan and type II collagen in response to damage or loss of cartilage matrix. This fundamental limitation of

chondrocytes is the core problem that has precluded the development of therapies that can prevent and cure osteoarthritis. Additionally, there is a need for a definitive diagnostic test for detecting early osteoarthritis, and a prognostic test that effectively monitors a patient's response to therapy.

5 Joint pain is the most common manifestation of early osteoarthritis. The pain tends to be episodic lasting days to weeks and remitting spontaneously. Although redness and swelling of joints is uncommon, joints become tender during a flare-up of osteoarthritis.

10 “Mild” or “early stage osteoarthritis” is difficult to diagnose. The physician relies primarily on the patient's history and physical exam to make the diagnosis of mild osteoarthritis. X-rays do not show the underlying early changes in articular cartilage. There are no recognized biochemical markers used to confirm the diagnosis of early stage osteoarthritis.

15 X-ray changes confirm the diagnosis of moderate osteoarthritis. X-rays of normal joints reveal well preserved symmetrical joint spaces. Changes seen on the x-rays of patients with osteoarthritis include new bone formation (osteophytes), joint space narrowing and sclerosis (bone thickening). There are no recognized biochemical markers used to confirm the diagnosis of “moderate osteoarthritis” at this stage.

20 The clinical exam of a joint with severe osteoarthritis reveals tenderness, joint deformity and a loss of mobility. Passive joint movement during examination may elicit crepitus or the grinding of bone-on-bone as the joint moves. X-ray changes are often profound: the joint space may be obliterated and misalignment of the joint can be seen. New bone formation (osteophytes) is prominent. Again, there are no recognized biochemical markers used to confirm the diagnosis of “severe osteoarthritis”.

25 “Osteoarthritis” is the most common chronic joint disease. It is characterized by progressive degeneration and eventual loss of cartilage. Currently, there is a need for an effective therapy that will alter the course of osteoarthritis. Further advances in preventing, modifying or curing the osteoarthritic disease process critically depends, at least in part, on a

thorough understanding of the molecular mechanisms underlying anabolic and catabolic processes in cartilage. Since cellular functions are substantially determined by the genes that the cells express, elucidating the genes expressed in articular cartilage at different developmental and disease stages will inevitably provide new insights into the molecules and 5 mechanisms involved in cartilage formation, injury, disease and repair.

cDNA libraries from putatively normal and severe osteoarthritic human cartilage tissue have been constructed (Kumar et al., 46th Annual Meeting, Orthopaedic Res. Soc., Abstract, p. 1031). However, this work does not adequately address the differentiation of chondrocyte gene expression from differing severities of osteoarthritic human cartilage (mild, 10 moderate, marked and severe). In addition, the “normal cartilage” samples were obtained from deceased donors more than 24 hours after death. Thus, this cDNA library does not truly reflect normal chondrocyte gene expression due to the rapid degeneration of RNA that occurs after cessation of perfusion to the sampled joint, as demonstrated by baboon studies, presented herein below.

15 **Summary of the Invention**

The invention relates to one or more profiles of gene expression for human fetal articular cartilage, and the cartilage of normal, mild, moderate, marked and severe osteoarthritic individuals, and thus to a method for identifying genes that play critical roles in cartilage injury, repair and disease progression. Given the inherently low anabolic activity in 20 adult chondrocytes, identification of key replicative and/or anabolic genes expressed by fetal but not adult chondrocytes, has important implications for developing novel disease modifying therapies for adult cartilage injury and osteoarthritis.

One aspect of the invention is to isolate chondrocyte enriched or chondrocyte-specific nucleic acid sequences.

25 In one embodiment, one or more nucleic acid sequences selected from the group consisting of the sequences identified in Figure 6A which correspond to genes 1-5807 identified in Figure 6 and/or sequences identified in Figure 13 are isolated.

In another embodiment, a vector comprising one or more nucleic acid sequences selected from the sequences identified in Figure 6A corresponding to genes 1-5807 of Figure 6 and/or sequences identified in Figure 13 are constructed.

In yet another embodiment, a host cell comprising the vector is constructed.

5 Another aspect of the invention is to provide a composition comprising one or more chondrocyte enriched or chondrocyte-specific nucleic acid sequences.

Another aspect of the invention is to provide a composition comprising one or more chondrocyte enriched or chondrocyte-specific nucleic acid sequences isolated from one or more of (a) a fetus, (b) normal, (c) mild osteoarthritic, (d) moderate osteoarthritic, (e) marked 10 osteoarthritic, or (f) severe osteoarthritic cartilage samples.

Another embodiment of the invention provides a composition comprising one or more nucleic acid sequences selected from the group of sequences identified in Figure 6B which are isolated from a fetal cDNA library as disclosed herein.

Another embodiment of the invention provides a composition comprising one or more nucleic acid sequences selected from the group of sequences identified in Figure 6C which are isolated from a normal cDNA library as disclosed herein.

Another embodiment of the invention provides a composition comprising one or more nucleic acid sequences selected from the group of sequences identified in Figure 6D which are isolated from a mild osteoarthritic chondrocyte library as disclosed herein.

20 Another embodiment of the invention provides a composition comprising one or more nucleic acid sequences selected from the group of sequences identified in Figure 6E which are isolated from a severe osteoarthritic chondrocyte library as disclosed herein.

Another embodiment of the invention provides a composition comprising one or more nucleic acid sequences selected from the group consisting of sequences as identified in 25 Figures 6B, 6C, 6D and 6E.

Another embodiment of the invention provides a composition comprising one or more nucleic acid sequences where at least one of the nucleic acid sequences is differentially expressed in cartilage from a patient diagnosed with mild osteoarthritis relative to cartilage from a normal individual, and where cartilage isolated from the normal individual is isolated
5 from cartilage tissue obtained less than 14 hours post-mortem.

Another embodiment of the invention provides a composition comprising one or more nucleic acid sequences where at least one of the nucleic acid sequences is differentially expressed in cartilage from a patient diagnosed with severe osteoarthritis relative to cartilage derived from a normal individual, where cartilage isolated from the normal individual is
10 isolated from cartilage tissue obtained less than 14 hours post-mortem.

15
16
17
18
19
20
21
22
23
24
25

Another embodiment of the invention provides a composition comprising one or more nucleic acid sequences where at least one of the nucleic acid sequences is differentially expressed in cartilage from a patient diagnosed with moderate osteoarthritis relative to cartilage from a normal individual, where cartilage isolated from the normal individual is
15 isolated from cartilage tissue obtained less than 14 hours post-mortem.

Another embodiment of the invention provides a composition comprising one or more nucleic acid sequences where at least one of the nucleic acid sequences is differentially expressed in cartilage from a patient diagnosed with marked osteoarthritis relative to cartilage derived from a normal individual, where cartilage isolated from the normal individual is
20 isolated from cartilage tissue obtained less than 14 hours post-mortem.

Another embodiment of the invention provides a composition comprising one or more nucleic acid sequences where at least one of the nucleic acid sequences is differentially expressed in cartilage isolated from any two or more of the following sources: (a) fetus, or (b) patient with mild osteoarthritis, (c) patient with moderate osteoarthritis, (d) patient with
25 marked osteoarthritis, (e) patient with severe osteoarthritis, and (f) cartilage isolated from cartilage tissue obtained from a normal individual.

Another embodiment of the invention provides a composition comprising one or more nucleic acids identified in Figure 9 and/or nucleic acid sequences identified in Figure 6A which correspond to the genes disclosed in Figure 9.

Another embodiment of the invention provides a composition comprising one or more 5 nucleic acids identified in Figure 11 and/or nucleic acid sequences identified in Figure 6A which correspond to the genes disclosed in Figure 11.

Another embodiment of the invention provides a composition comprising one or more nucleic acids identified in Figure 6A which correspond to the genes disclosed in Figures 15 and Figures 16.

Another embodiment of the invention provides a composition comprising one or more 10 nucleic acids identified in Figure 6A which correspond to the genes disclosed in Figure 6.

Another embodiment of the invention provides a composition comprising one or more nucleic acid sequences selected from sequences identified in Figure 13.

A further aspect of the invention relates to nucleic acid arrays comprising a plurality 15 of chondrocyte enriched or chondrocyte-specific nucleic acid member sequences.

In one embodiment, the invention provides an array comprising a plurality of nucleic acid members, where at least one member is differentially expressed in cartilage from a patient diagnosed with mild osteoarthritis, as compared to cartilage from a normal individual, and a solid substrate, where each nucleic acid member has a unique position on the array and 20 is stably associated with the solid substrate.

In another embodiment, the invention provides an array comprising a plurality of nucleic acid members, where at least one member is differentially expressed in cartilage isolated from a patient diagnosed with severe osteoarthritis, as compared to cartilage from a normal individual, and a solid substrate, where each nucleic acid member has a unique 25 position on the array and is stably associated with the solid substrate.

In another embodiment, the invention provides an array comprising a plurality of nucleic acid members, where at least one member is differentially expressed in cartilage isolated from a patient diagnosed with moderate osteoarthritis, as compared to cartilage from a normal individual, and a solid substrate, where each nucleic acid member has a unique 5 position on the array and is stably associated with the solid substrate.

In another embodiment, the invention provides an array comprising a plurality of nucleic acid members, where at least one member is differentially expressed in cartilage isolated from a patient diagnosed with marked osteoarthritis, as compared to cartilage from a normal individual, and a solid substrate, where each nucleic acid member has a unique 10 position on the array and is stably associated with the solid substrate.

In another embodiment, the invention provides an array comprising a plurality of nucleic acid members, where at least one member is differentially expressed in cartilage isolated from a fetus, as compared to cartilage from a normal individual, and a solid substrate, where each nucleic acid member has a unique position on the array and is stably associated 15 with the solid substrate.

In a preferred embodiment, cartilage is isolated from a living normal individual.

In another preferred embodiment, the cartilage is isolated from the normal individual in less than 14 hours post-mortem.

In another embodiment, the invention provides an array comprising a plurality of 20 nucleic acid members and a solid substrate, where at least one member is differentially expressed in cartilage isolated from any two or more of the following sources: (a) fetus, (b) patient with mild osteoarthritis, (c) patient with moderate osteoarthritis, (d) patient with marked osteoarthritis, (e) patient with severe osteoarthritis or (f) cartilage isolated from normal individual, and where each nucleic acid member has a unique position on the array 25 and is stably associated with the solid substrate.

In one embodiment, each nucleic acid member on an array according to the invention, is at least 50 nucleotides.

In another embodiment, an array according to the invention comprises from 10 to 20,000 positions.

5 In yet another embodiment, an array according to the invention further includes negative and positive control sequences and RNA quality control sequences. Control sequences can be selected from the group consisting of cDNA sequences of housekeeping genes, plant gene sequences (and/or their cDNA sequences), bacterial sequences, PCR products, vector sequences, and combinations thereof.

10 Another aspect of the invention relates to novel methods for diagnosing osteoarthritis.

In one embodiment, a method for diagnosing mild osteoarthritis in a patient comprises hybridizing a nucleic acid sample corresponding to RNA (e.g., a sample comprising RNA or cDNA or amplified products of RNA or cDNA) to an array comprising a solid substrate and a plurality of nucleic acid members, where at least one member is differentially expressed in cartilage isolated from a patient diagnosed with mild osteoarthritis, as compared to cartilage isolated from a normal individual and where each nucleic acid member has a unique position and is stably associated with the solid substrate. The cartilage isolated from the normal individual is isolated from cartilage tissue less than 14 hours post-mortem. Hybridization of the nucleic acid sample to one or more of the nucleic acid members is indicative of mild osteoarthritis.

20 In another embodiment, a method of diagnosing moderate osteoarthritis in a patient comprises hybridizing a nucleic acid sample corresponding to RNA to an array comprising a solid substrate and a plurality of nucleic acid members, where at least one member is differentially expressed in cartilage isolated from a patient diagnosed with moderate osteoarthritis, as compared to cartilage isolated from a normal individual and where each nucleic acid member has a unique position and is stably associated with the solid substrate. Cartilage isolated from the normal individual is isolated from cartilage tissue less than 14

hours post-mortem. Hybridization of the nucleic acid sample to one or more of the nucleic acid members on the array is indicative of moderate osteoarthritis.

In yet another embodiment, a method of diagnosing marked osteoarthritis in a patient comprises hybridizing a nucleic acid sample corresponding to RNA to an array comprising a solid substrate and a plurality of nucleic acid members, where at least one member is differentially expressed in cartilage isolated from a patient diagnosed with marked osteoarthritis, as compared to cartilage isolated from a normal individual and each nucleic acid member has a unique position and is stably associated with the solid substrate. Like the above arrays, cartilage isolated from the normal individual is isolated from cartilage tissue less than 14 hours post-mortem. Hybridization of the nucleic acid sample to one or more of the nucleic acid members is indicative of marked osteoarthritis.

In a further embodiment, a method of diagnosing severe osteoarthritis in a patient comprises hybridizing a nucleic acid sample corresponding to RNA to an array comprising a solid substrate and a plurality of nucleic acid members, where at least one member is differentially expressed in cartilage isolated from a patient diagnosed with severe osteoarthritis, as compared to cartilage isolated from a normal individual and each nucleic acid member has a unique position and is stably associated with the solid substrate. Like the above arrays, cartilage from the normal individual is isolated from cartilage tissue less than 14 hours post-mortem. Hybridization of the nucleic acid sample to one or more of the nucleic acid members is indicative of severe osteoarthritis.

In a preferred embodiment, the method of diagnosis comprises isolating a cartilage sample from a patient at a specific stage of osteoarthritis (e.g., mild, moderate, marked, or severe).

In another preferred embodiment, the method of diagnosis further comprises the step of preparing an RNA sample from a cartilage sample.

In another preferred embodiment, the method of diagnosis further comprises the step of preparing an RNA sample from blood.

In another preferred embodiment, the method of diagnosis further comprises the step of preparing an RNA sample from synovial fluid.

Another aspect of the invention relates to a method of identifying an agent that increases or decreases the expression of one or more nucleic acid sequences that are 5 differentially expressed in a chondrocyte derived from a fetus or from patient(s) with a chondrocyte disease selected from the group consisting of: mild osteoarthritis, moderate osteoarthritis, marked osteoarthritis and severe osteoarthritis. The method comprises incubating a chondrocyte isolated from a cartilage sample obtained from a normal individual less than 14 hours post-mortem with a candidate agent. RNA is isolated from the 10 chondrocyte and a probe is hybridized to the RNA which corresponds to a nucleic acid sequence which is differentially expressed in a chondrocyte from any two or more of the following developmental or disease stages: a fetus, normal, mild osteoarthritic, moderate osteoarthritic, marked osteoarthritic and severe osteoarthritic. Differential hybridization of the probe to the RNA from normal individual(s) relative to hybridization of the probe to RNA 15 from any one or more of: fetus(es), patient(s) with mild osteoarthritis, patient(s) with moderate osteoarthritis, patient(s) with marked osteoarthritis and patient(s) with severe osteoarthritis identifies the RNA which specifically hybridizes to the probe as a differentially expressed chondrocyte-specific nucleic acid sequence and identifies the candidate agent as one which increases or decreases the expression of the chondrocyte-specific nucleic acid 20 sequence.

The method also can be performed by evaluating cDNA corresponding to RNAs obtained from chondrocytes.

This method also can be performed by evaluating cDNA corresponding to RNAs obtained from blood.

25 This method also can be performed by evaluating cDNA corresponding to RNAs obtained from synovial fluid.

The invention further relates to methods of preparing chondrocyte cDNA libraries.

5 In one embodiment, a method of preparing a chondrocyte cDNA library comprises: a) isolating chondrocytes from a cartilage sample from a normal individual, where the cartilage sample is obtained less than 14 hours post-mortem, b) isolating total RNA from the chondrocytes, c) synthesizing cDNA from the mRNA in the total RNA, and d) ligating the cDNA into a vector.

In another embodiment, a method of preparing a chondrocyte cDNA library comprises: a) isolating chondrocytes from a cartilage sample from a normal individual, wherein the normal individual is living, b) isolating total RNA from the chondrocytes, c) synthesizing cDNA from mRNA in the total RNA, and d) ligating the cDNA into a vector.

10 In another embodiment, a method of preparing a chondrocyte cDNA library comprises: a) isolating chondrocytes from a cartilage sample from a fetus, b) isolating total RNA from the chondrocytes, c) synthesizing cDNA from mRNA in the total RNA, and d) ligating the cDNA into a vector.

15 In another embodiment, a method of preparing a chondrocyte cDNA library is provided comprising, a) isolating chondrocytes from a cartilage sample from a patient diagnosed with mild, moderate, marked or severe osteoarthritis, b) isolating total RNA from the chondrocytes, c) synthesizing cDNA from mRNA in the total RNA, and d) ligating the cDNA into a vector.

20 The invention also relates to a method of making an array which comprises a plurality of nucleic acid members comprising nucleic acid sequences selected from the group consisting of sequences of Figure 14 on a solid support comprising a surface with a plurality of pre-selected unique regions. The method comprises: spotting each nucleic acid member individually onto a unique pre-selected region, and stably associating each nucleic acid member with the solid support at the pre-selected region.

25 In a preferred embodiment, at least one nucleic acid member is differentially expressed in cartilage isolated from a patient diagnosed with mild, moderate, marked, or severe osteoarthritis, as compared to cartilage isolated from a normal individual, where

cartilage isolated from the normal individual is isolated from cartilage tissue less than 14 hours post-mortem.

In another preferred embodiment, at least one nucleic acid member is differentially expressed in cartilage isolated from a patient diagnosed with mild, moderate, marked, or 5 severe osteoarthritis, as compared to cartilage isolated from a fetus.

In another preferred embodiment, at least one nucleic acid member is differentially expressed in cartilage isolated from a fetus, as compared to a cartilage isolated from a normal individual, where cartilage isolated from the normal individual is isolated from cartilage tissue less than 14 hours post-mortem.

10 In another preferred embodiment, at least one nucleic acid member is differentially expressed in cartilage isolated from any two of the following sources: (a) fetus, (b) a normal individual where cartilage isolated from the normal individual is isolated from cartilage tissue less than 14 hours post-mortem, and (c) a patient diagnosed with mild osteoarthritis, (d) a patient diagnosed with moderate osteoarthritis, (e) a patient diagnosed with marked osteoarthritis, or (f) a patient diagnosed with severe osteoarthritis.

15 The invention also provides kits comprising one or more of the compositions and/or arrays described above and packaging means therefore.

Brief Description of the Drawings

20 The objects and features of the invention can be better understood with reference to the following detailed description and drawings.

Figure 1, is a graph according to one embodiment of the invention showing the relative EST frequency level of selected extracellular matrix (ECM) proteins among the fetal, normal, mild osteoarthritic and severe osteoarthritic cartilage cDNA libraries. The percentages were calculated by dividing the number of ESTs matched to a certain type of 25 ECM protein by the total number of ECM. ESTs per library. Legend: COL=collagen,

PGL=proteoglycan, CMP=cartilage matrix proteins, OSN=osteonectin, FN=fibronectin, CRTL 1=cartilage link protein.

Figure 2, is a graph according to one embodiment of the invention showing the relative EST frequency level of collagens among the fetal, normal, mild osteoarthritic and severe osteoarthritic cartilage cDNA libraries. The percentages were calculated by dividing the total number of collagen ESTs in a particular library by the total number of collagen ESTs from the four cartilage libraries.

Figure 3, is a graph according to one embodiment of the invention showing the relative EST levels of specific collagen types among the fetal, normal, mild osteoarthritic and severe osteoarthritic cartilage cDNA libraries. Percentages were calculated by dividing the total number of ESTs for each type of collagen in a particular library by the total number of collagen ESTs from each library.

25

Figure 5, is a table according to one embodiment of the invention showing the total number of ESTs in each of the four cDNA libraries and the breakdown of what the ESTs represent, including the number of novel sequences (ie. ESTs with no significant match) in each library.

Figure 5A, is a graphical representation of the data presented in Figure 5 according to one embodiment of the invention.

Figure 6, is a table listing the unique known genes (5,807) identified in the four cDNA libraries to date according to one embodiment of the invention.

Figure 6A, is a table listing the names of the EST sequences identified in the four cDNA libraries that represent each of the unique known genes identified in Figure 6 according to one embodiment of the invention.

Figure 6B is a table listing the names of all of the EST sequences identified from the cDNA library constructed from fetal cartilage tissue according to one embodiment of the invention.

Figure 6C is a table listing the names of all of the EST sequences identified from the cDNA library constructed from normal cartilage tissue where such tissue is obtained less than 14 hours post-mortem according to one embodiment of the invention.

Figure 6D is a table listing the names of all of the EST sequences identified from the cDNA library constructed from cartilage of patients with mild osteoarthritis according to one embodiment of the invention.

Figure 6E is a table listing the names of all of the EST sequences identified from the cDNA library constructed from cartilage of patients with severe osteoarthritis according to one embodiment of the invention.

Figure 7, is a table showing the characterization of the total number of ESTs from the four cDNA libraries (57,422) based on the functional classification of unique known genes represented by the ESTs according to one embodiment of the invention.

Figure 7A, is a graphical representation of the data presented in Figure 7 according to one embodiment of the invention.

Figure 8, is a list of known and novel EST clones from the mild and severe cDNA libraries comprising a microarray according to one embodiment of the invention.

Figure 9, is a table showing candidate upregulated genes detected in the mild osteoarthritis cDNA library based on the microarray analysis according to one embodiment of the invention.

5 Figure 10, is a table showing candidate down-regulated genes detected in the mild osteoarthritis cDNA library based on the microarray analysis according to one embodiment of the invention.

Figure 11, is a table showing candidate up-regulated genes detected in the severe osteoarthritis cDNA library based on the microarray analysis according to one embodiment of the invention.

10 Figure 12, is a table showing candidate down-regulated genes detected in the severe osteoarthritic cDNA library based on the microarray analysis according to one embodiment of the invention.

15 Figure 13, is a table listing the EST sequence names representing novel sequences identified in each of the four cDNA libraries to date according to one embodiment of the invention.

Figure 14, is a CD ROM, attached hereto, containing all of the EST sequences identified from the four human cartilage cDNA libraries according to one embodiment of the invention. A subset of sequences in this figure was filed as a CD ROM in U.S. provisional application 60/271,955, filed February 28, 2001 and as a hard copy in U.S. provisional application 20 60/275,017, filed March 12, 2001. The names of all of the EST sequences on the CD-ROM are listed in Figures 6B, 6C, 6D and 6E.

25 Figure 14A, is a hard copy of a subset of the EST sequences identified in Figure 14 which represents the sequences not filed as a CD ROM in U.S. provisional application 60/271,955, filed February 28, 2001 or as a hard copy in U.S. provisional application 60/275,017, filed March 12, 2001. A CD ROM corresponding to Figure 14A is attached hereto.

Figure 15, contains a list of genes that have been identified through EST frequency analysis as being differentially expressed between fetal and normal cDNA libraries according to one embodiment of the invention.

Figure 16, contains a list of genes that have been identified through EST frequency analysis as being differentially expressed between mild and severe osteoarthritis cDNA libraries according to one embodiment of the invention.

Figure 17, is a bar graph showing the level of beta-2 microglobulin (B2M) in synovial fluid from normal individuals and patients with different stages of osteoarthritis according to one embodiment of the invention. Legend: nor=normal individual, miao=patient with mild osteoarthritis, mooa=patient with moderate osteoarthritis, maoa=patient with marked osteoarthritis, seoa=patient with severe osteoarthritis.

Figure 18, is a bar graph showing the level of beta 2 microglobulin (B2M) in medium cultured from cartilage from patients with severe osteoarthritis at varying time periods during culturing according to one embodiment of the invention.

Figure 19, is a black and white representation of a two-color fluorescent scan according to one embodiment of the invention showing genes preferentially expressed in non-B2M-treated chondrocytes and genes preferentially expressed in B2M-treated chondrocytes. B2M=beta 2 microglobulin.

Detailed Description of the Invention

The invention relates to methods of profiling gene sequences expressed in human chondrocytes to identify differential gene expression in chondrocytes at different stages of development and disease. Differentially expressed genes and their products (e.g., mRNAs and proteins) can be used in methods for diagnosis, prognosis, screening, or treatment of osteoarthritis.

25 Definitions

The following definitions are provided for specific terms which are used in the following written description.

As used herein, "osteoarthritis" refers to a chronic disease in which the articular cartilage that lies on the ends of bones that form the articulating surface of the joints gradually degenerates over time. Cartilage degeneration can be caused by an imbalanced catabolic activity (removal of "old" cells and matrix components) and anabolic activity (production of "new" cells and molecules) (Westacott et al., 1996, *Semin Arthritis Rheum*, 25:254-72).

As used herein, "cartilage" or "articular cartilage" refers to elastic, translucent connective tissue in mammals, including human and other species. Cartilage is composed predominantly of chondrocytes, type II collagen, small amounts of other collagen types, other noncollagenous proteins, proteoglycans and water, and is usually surrounded by a perichondrium, made up of fibroblasts, in a matrix of type I and type II collagen as well as other proteoglycans. Although most cartilage becomes bone upon maturation, some cartilage remains in its original form in locations such as the nose, ears, knees, and other joints. The cartilage has no blood or nerve supply and chondrocytes are the only type of cell in this tissue.

As used herein, "chondrocyte" refers to cartilage cells.

As used herein, "synovial fluid" refers to fluid secreted from the "synovial sac" which surrounds each joint. Synovial fluid serves to protect the joint, lubricate the joint and provide nourishment to the articular cartilage. Synovial fluid useful according to the invention contains cells from which RNA can be isolated according to methods well known in the art as described herein.

As used herein, the term "osteoarthritis (OA) staging" or "osteoarthritis (OA) grading" refers to determining the degree of advancement or progression of the disease in the cartilage. In order to classify cartilage into different disease stages, a scoring system is used according to known methods in the art. Preferably the scoring system described in Marshall (Marshall W., 1996, *The Journal of Rheumatology*, 23:582-584, incorporated by reference) is used. According to this method, each of the 6 articular surfaces (patella, femoral trochlea, medial

femoral condyle, medial tibial plateau, lateral femoral condyle and lateral tibial plateau) is assigned a cartilage grade based on the worst lesion present on that specific surface. A scoring system is then applied in which each articular surface receives an OA severity number value that reflects the cartilage severity grade for that surface. For example, if the medial 5 femoral condyle has a grade I lesion as its most severe cartilage damage a value of 1 is assigned. A total score for the patient is then derived from the sum of the scores on the 6 articular surfaces. Based on the total score, each patient is placed into one of 4 OA groups: mild (early) (1-6), moderate (7-12), marked (13-18) and severe (>18).

As used herein, "diagnosis" refers to a process of determining if an individual is 10 afflicted with a disease or ailment. "Diagnosis of OA" or "OA diagnosis", according to the invention, means determining if an individual is afflicted with OA, or, once a patient is diagnosed, determining the OA stage or grade as used herein based on the medical history and physical examination of the patient using methods known in the art (i.e., joint X ray). Preferably, OA stages are measured using the scoring system described by Marshall, *supra*. 15 "Prognosis of OA" refers to a prediction of the probable occurrence and/or progression of OA in a patient, as well as the likelihood of recovery from OA, or the likelihood of ameliorating symptoms of OA or the likelihood of reversing the effects of OA.

As used herein, "patient" refers to a mammal who is diagnosed with a mild, moderate, 20 marked, or severe form of OA.

As used herein, "normal" refers to an individual who has not shown any OA symptoms or has not been diagnosed with cartilage injury or OA. "Normal", according to the invention, also refers to a sample taken from a normal individual within 14 hours post-mortem. A normal cartilage tissue sample, for example, refers to the whole or a piece of cartilage isolated from cartilage tissue within 14 hours post-mortem from an individual who 25 was not diagnosed with OA and whose corpse does not show any symptoms of OA at the time of tissue removal. In alternative embodiments of the invention, the "normal" cartilage tissue sample is isolated from cartilage tissue less than 14 hours post-mortem, e.g., within 13 hours, 12 hours, 11 hours, 10 hours, 9 hours, 8 hours, 7 hours, 6 hours, 5 hours, 4 hours, 3

hours, 2 hours, or 1 hour post-mortem. In one embodiment of the invention, the "normal" cartilage sample is isolated at 14 hours post-mortem and the integrity of mRNA samples extracted is confirmed.

As used herein, "mRNA integrity" refers to the quality of mRNA extracts from 5 cartilage samples. mRNA extracts with good integrity do not appear to be degraded when examined by methods well known in the art, for example, by RNA agarose gel electrophoresis (e.g., Ausubel et al., John Weley & Sons, Inc., 1997, *Current Protocols in Molecular Biology*). Preferably, the mRNA samples have good integrity (e.g., less than 10%, preferably, less than 5%, and more preferably, less than 1% of the mRNA is degraded) to 10 truly represent the gene expression levels of the cartilage samples from which they are extracted.

As used herein, "fetal" cartilage samples refer to samples taken from a fetus. The chondrocytes of fetal cartilage have a higher level of metabolic activity and cell division rates as compared to chondrocytes from cartilage derived from either a normal adult or from an 15 adult diagnosed with any stage of OA (mild, moderate, marked and severe).

As used herein, "nucleic acid(s)" generally refers to any polyribonucleotide or poly-deoxyribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. "Nucleic acids" include, without limitation, single-and double-stranded nucleic acids. As used herein, the term "nucleic acid(s)" also includes DNAs or RNAs as described above, that 20 contain one or more modified bases. Thus, DNAs or RNAs with backbones modified for stability or for other reasons are "nucleic acids". The term "nucleic acids" as it is used herein embraces such chemically, enzymatically or metabolically modified forms of nucleic acids, as well as the chemical forms of DNA and RNA characteristic of viruses and cells, including for example, simple and complex cells. A "nucleic acid" or "nucleic acid sequence" may also be 25 an expressed sequence tag (EST) according to some embodiments of the invention. An EST is a small part of the expressed sequence of a gene (i.e., the "tag" of a sequence), made from cDNA. An EST can be used to fish the rest of the gene out of the chromosome, by matching base pairs with part of the expressed sequence of the gene.

As used herein, “isolated” or “purified” when used in reference to a nucleic acid means that a naturally occurring sequence has been removed from its normal cellular (e.g., chromosomal) environment or is synthesized in a non-natural environment (e.g., artificially synthesized). Thus, an “isolated” or “purified” sequence may be in a cell-free solution or 5 placed in a different cellular environment. The term “purified” does not imply that the sequence is the only nucleotide present, but that it is essentially free (about 90-95% pure) of non-nucleotide material naturally associated with it, and thus is distinguished from isolated chromosomes.

As used herein, the term “probe” refers to an oligonucleotide which forms a duplex 10 structure with a sequence in the target nucleic acid, due to complementarity of at least one sequence in the probe with a sequence in the target region.

As defined herein, a “nucleic acid array” refers a plurality of unique nucleic acids (or “nucleic acid members”) attached to one surface of a solid support at a density exceeding 20 different nucleic acids/cm² where each of the nucleic acid members is attached to the surface 15 of the solid support in a non-identical pre-selected region. In one embodiment, the nucleic acid member attached to the surface of the solid support is DNA. In a preferred embodiment, the nucleic acid member attached to the surface of the solid support is cDNA. In another preferred embodiment, the nucleic acid member attached to the surface of the solid support is cDNA synthesized by polymerase chain reaction (PCR). Preferably, a nucleic acid member 20 of the array according to the invention is at least 50 nucleotides in length. Preferably, a nucleic acid member of the array is less than 6,000 nucleotides in length. More preferably, a nucleic acid member of the array comprises an array less than 500 nucleotides in length. In one embodiment, the array comprises at least 500 different nucleic acid members attached to 25 one surface of the solid support. In another embodiment, the array comprises at least 10 different nucleic acid members attached to one surface of the solid support. In yet another embodiment, the array comprises at least 10,000 different nucleic acid members attached to one surface of the solid support. In yet another embodiment, the array comprises at least

15,000 different nucleic acid members attached to one surface of the solid support. The term "nucleic acid", as used herein, is interchangeable with the term "nucleic acid".

As used herein, "a plurality of" or "a set of" refers to more than two, for example, 3 or more, 100 or more, or 1000 or more.

5 As used herein, "attaching" or "spotting" refers to a process of depositing a nucleic acid onto a solid substrate to form a nucleic acid array such that the nucleic acid is irreversibly bound to the solid substrate via covalent bonds, hydrogen bonds or ionic interactions.

As used herein, "stably associated" refers to a nucleic acid that is irreversibly bound to a solid substrate to form an array via covalent bonds, hydrogen bonds or ionic interactions such that the nucleic acid retains its unique pre-selected position relative to all other nucleic acids that are stably associated with an array, or to all other pre-selected regions on the solid substrate under conditions in which an array is typically analyzed (i.e., during one or more steps of hybridization, washes, and/or scanning, etc.).

10 As used herein, "solid substrate" or "solid support" refers to a material having a rigid or semi-rigid surface. The terms "substrate" and "support" are used interchangeably herein with the terms "solid substrate" and "solid support". The solid support may be biological, non-biological, organic, inorganic, or a combination of any of these, existing as particles, strands, precipitates, gels, sheets, tubing, spheres, beads, containers, capillaries, pads, slices, 15 films, plates, slides, chips, etc. Often, the substrate is a silicon or glass surface, (poly)tetrafluoroethylene, (poly)vinylidendifluoride, polystyrene, polycarbonate, a charged membrane, such as nylon 66 or nitrocellulose, or combinations thereof. In a preferred embodiment, the solid support is glass. Preferably, at least one surface of the substrate will be substantially flat. Preferably, the surface of the solid support will contain reactive groups, 20 including, but not limited to, carboxyl, amino, hydroxyl, thiol, and the like. In one embodiment, the surface is optically transparent.

As used herein, “pre-selected region”, “predefined region”, or “unique position” refers to a localized area on a substrate which is, was, or is intended to be used for the deposit of a nucleic acid and is otherwise referred to herein in the alternative as a “selected region” or simply a “region.” The pre-selected region may have any convenient shape, e.g., circular, 5 rectangular, elliptical, wedge-shaped, etc. In some embodiments, a pre-selected region is smaller than about 1 cm², more preferably less than 1 mm², still more preferably less than 0.5 mm², and in some embodiments less than 0.1 mm². A nucleic acid member at a “pre-selected region”, “predefined region”, or “unique position” is one whose identity (e.g., sequence) can be determined by virtue of its position at the region or unique position.

10 As used herein, a “nucleic acid target” or “a target nucleic acid” is defined as a nucleic acid capable of binding to a nucleic acid member of complementary sequence through one or more types of chemical bonds, usually through complementary base pairing, i.e., through hydrogen bond formation. As used herein, a nucleic acid target may include natural (i. e., A, G, C, or T) or modified bases (7-deazaguanosine, inosine, etc.). In addition, the 15 bases in nucleic acid probe may be joined by a linkage other than a phosphodiester bond, so long as it does not interfere with hybridization (i.e., the probe still specifically binds to its complementary sequence under standard stringent or selective hybridization conditions). Thus, nucleic acid targets may be peptide nucleic acids in which the constituent bases are joined by peptide bonds rather than phosphodiester linkages. Preferably, the nucleic acid 20 targets are derived from human cartilage, blood or synovial fluid extracts. More preferably, the nucleic acid targets are single- or double-stranded DNA, RNA, or DNA-RNA hybrids, from human cartilage, blood or synovial fluid RNA extracts, and preferably from mRNA extracts.

25 As used herein, a “cartilage nucleic acid sample”, refers to nucleic acids derived from cartilage. Preferably, a cartilage nucleic acid sample is RNA or is a nucleic acid corresponding to RNA, for example, cDNA.

As used herein, the term “hybridizing to” or “hybridization” refers to the hydrogen binding with a complementary nucleic acid, via an interaction between for example, a target nucleic acid sequence and a nucleic acid member in an array.

As used herein, “specific hybridization” or “selective hybridization” refers to hybridization which occurs when two nucleic acid sequences are substantially complementary (at least about 65% complementary over a stretch of at least 14 to 25 nucleotides, preferably at least about 75%, more preferably at least about 90% complementary). See Kanehisa, M., 1984, *Nucleic acids Res.*, 12:203, incorporated herein by reference. As a result, it is expected that a certain degree of mismatch is tolerated. Such mismatch may be small, such as a mono-, di- or tri-nucleotide. Alternatively, a region of mismatch can encompass loops, which are defined as regions in which there exists a mismatch in an uninterrupted series of four or more nucleotides. Numerous factors influence the efficiency and selectivity of hybridization of two nucleic acids, for example, a nucleic acid member on a array, to a target nucleic acid sequence. These factors include nucleic acid member length, nucleotide sequence and/or composition, hybridization temperature, buffer composition and potential for steric hindrance in the region to which the nucleic acid member is required to hybridize. A positive correlation exists between the nucleic acid member length and both the efficiency and accuracy with which a nucleic acid member will anneal to a target sequence. In particular, longer sequences have a higher melting temperature (T_M) than do shorter ones, and are less likely to be repeated within a given target sequence, thereby minimizing promiscuous hybridization. Hybridization temperature varies inversely with nucleic acid member annealing efficiency, as does the concentration of organic solvents, e.g., formamide, that might be included in a hybridization mixture, while increases in salt concentration facilitate binding. Under stringent annealing conditions, longer nucleic acids, hybridize more efficiently than do shorter ones, which are sufficient under more permissive conditions.

As used herein, the term “differential hybridization” refers to a situation where a probe hybridizes to the same nucleic acid sequence obtained from two or more samples at different levels. “Differential hybridization” means that the ratio of the level of hybridization

of the probe to the nucleic acid sequence isolated from one sample as compared to the nucleic acid sequence isolated from another sample is not equal to 1.0. For example, the ratio of the level of hybridization of the probe to the nucleic acid sequence isolated from one sample as compared to the nucleic acid sequence isolated from another sample is greater than or less 5 than 1.0 and includes greater than 1.2 and less than 0.7, greater than 1.5 and less than 0.5. A differential hybridization also exists if the hybridization is detectable in one sample but not another sample.

As herein used, the term “standard stringent conditions” means hybridization will occur only if there is at least 95% and preferably, at least 97% identity between the sequences, 10 where the region of identity comprises at least 10 nucleotides. In one embodiment, the sequences hybridize under stringent conditions following incubation of the sequences overnight at 42°C, followed by stringent washes (0.2X SSC at 65° C). As several factors affect the stringency of hybridization, the combination of parameters is more important than the absolute measure of a single factor.

15 As used herein, the term “level of expression” refers to the measurable expression level of a given nucleic acid. The level of expression of a nucleic acid is determined by methods well known in the art. The term “differentially expressed” or “changes in the level of expression” refers to an increase or decrease in the measurable expression level of a given nucleic acid. As used herein, “differentially expressed” when referring to microarray analysis 20 means the ratio of the expression level of a given nucleic acid in one sample and the expression level of the given nucleic acid in another sample is not equal to 1.0. “Differentially expressed” when referring to microarray analysis according to the invention also means the ratio of the expression level of a given nucleic acid in one sample and the expression level of the given nucleic acid in another sample is greater than or less than 1.0 25 and includes greater than 1.2 and less than 0.7, as well as greater than 1.5 and less than 0.5. A nucleic acid also is said to be differentially expressed in two samples if one of the two samples contains no detectable expression of the nucleic acid. Absolute quantification of the level of expression of a nucleic acid can be accomplished by including known

concentration(s) of one or more control nucleic acid species, generating a standard curve based on the amount of the control nucleic acid and extrapolating the expression level of the “unknown” nucleic acid species from the hybridization intensities of the unknown with respect to the standard curve. The level of expression is measured by hybridization analysis
5 using labeled target nucleic acids according to methods well known in the art. The label on the target nucleic acid can be a luminescent label, an enzymatic label, a radioactive label, a chemical label or a physical label. Preferably, target nucleic acids are labeled with a fluorescent molecule. Preferred fluorescent labels include, but are not limited to, fluorescein, amino coumarin acetic acid, tetramethylrhodamine isothiocyanate (TRITC), Texas Red, Cy3
10 and Cy5.

As used herein “differentially expressed” when referring to EST analysis refers to the relative expression level of a gene based on the frequency of ESTs representing the gene derived from a cDNA library as compared to the frequency of ESTs representing the same gene derived from another cDNA library. As described herein, the “relative EST frequency”
15 of an EST is calculated by dividing the number of ESTs representing each specific gene by the total number of ESTs analyzed. Differences in “relative EST frequency” may be used as an indication of differential gene expression.

As used herein, the term “significant match”, when referring to nucleic acid sequences, means that two nucleic acid sequences exhibit at least 65% identity, at least 70%,
20 at least 75%, at least 80%, at least 85%, and preferably, at least 90% identity, using comparison methods well known in the art (i.e., Altschul, S.F. et al., 1997, *Nucl. Acids Res.*, 25:3389-3402; Schäffer, A.A. et al., 1999, *Bioinformatics* 15:1000-1011). As used herein, “significant match” encompasses non-contiguous or scattered identical nucleotides so long as the sequences exhibit at least 65%, and preferably, at least 70%, at least 75%, at least 80%, at
25 least 85%, and preferably, at least 90% identity, when maximally aligned using alignment methods routine in the art.

As used herein, a “novel sequence” or “novel expressed sequence tag (EST)” refers to a nucleic acid sequence which has no significant match to any existing sequence in the “nt”,

“nr”, “est”, “gss” and “htg” databases available through NCBI at the time each novel sequence was compared. “No significant match” preferably refers to a less than 65% match between a novel sequence being queried against other sequences in the database, and preferably, a less than 50% match, a less than 40% match, or a less than 30% match, after 5 maximally aligning sequences using methods routine in the art.

As used herein, a “known sequence” refers to a nucleic acid sequence which has significant match to at least one existing sequence in the “nt”, “nr”, “est”, “gss” and “htg” databases available through NCBI. “Known sequence with a function” refers to a nucleic acid with significant match to an existing sequence which encodes a polypeptide with a 10 known function. “Known sequence with no function” refers to a nucleic acid that exhibits a significant match to an existing sequence which encodes a polypeptide of unknown function.

As used herein, a “chondrocyte-specific nucleic acid” is a nucleic acid sequence which is expressed at a detectable level in a chondrocyte and is not expressed at a detectable level in any other cell types as indicated by having no significant match to any sequence in any of the 15 available databases comprising sequences from other cell types.

As used herein, a “chondrocyte enriched nucleic acid” or “chondrocyte enriched sequence” refers to a sequence which is differentially expressed in chondrocytes as compared to non-chondrocytes.

As used herein, “indicative of disease” refers to an expression pattern which is 20 diagnostic of disease such that the expression pattern is found significantly more often in patients with a disease than in patients without the disease (as determined using routine statistical methods setting confidence levels at 95%). Preferably, an expression pattern which is indicative of disease is found in at least 70% of patients who have the disease and is found in less than 10% of patients who do not have the disease. More preferably, an expression 25 pattern which is indicative of disease is found in at least 75%, at least 80%, at least 85%, at least 90%, at least 95% or more in patients who have the disease and is found in less than

10%, less than 8%, less than 5%, less than 2.5%, or less than 1% of patients who do not have the disease.

As used herein, "gene expression pattern" or "gene expression profile" comprises the pattern (i.e., qualitatively and/or quantitatively) of expression of one or more expressed 5 nucleic acid sequences where one or more members of the set are differentially expressed.

As used herein, "a nucleic acid array expression profile" is generated from the hybridization of nucleic acids derived from a sample to one or more nucleic acid members comprising an array according to the invention.

As used herein, a "therapeutic agent" or "agent" refers to a compound that increases or 10 decreases the expression of a nucleic acid sequence that is differentially expressed in a chondrocyte from any two of the following developmental or osteoarthritis disease stages: (a) fetal, (b) mild, (c) moderate, (d) marked and (e) severe, or (f) chondrocyte from a normal individual, as defined herein. A therapeutic agent according to the invention also refers to a compound that increases or decreases the anabolic activity of a chondrocyte. The invention 15 provides for a "therapeutic agent" that 1) prevents the onset of osteoarthritis; 2) reduces, delays, or eliminates osteoarthritis symptoms such as pain, swelling, weakness and loss of functional ability in the afflicted joints; 3) reduces, delays, or eliminates cartilage degeneration, and/or enhances chondrocyte metabolic activity and cell division rates; and/or 4) restores one or more expression profiles of one or more disease-indicative nucleic acids of 20 a patient to a profile more similar to that of a normal individual when administered to a patient.

Identifying Chondrocyte Enriched and Chondrocyte-Specific Nucleic acid Sequences

cDNA libraries were constructed from human fetal, normal, mild osteoarthritic and severe osteoarthritic cartilage samples. The known and novel clones derived from these 25 libraries were then used to construct human chondrocyte-specific microarrays to generate differential gene expression profiles useful as a diagnostic tool for detection of mild (early

stage) osteoarthritis. Arrays of the invention are useful as a gold standard for osteoarthritis diagnosis and for use to identify and monitor therapeutic efficacy of new drug targets.

One effective and rapid way of characterizing gene expression patterns in a given tissue is through large-scale partial sequencing of a cDNA library produced from such a tissue
5 to generate expressed sequence tags (ESTs). This approach has provided both quantitative and qualitative information on gene expression in a variety of tissues and cells (4-7). Since cDNA libraries represent gene transcription in the cells of the tissue used to construct the library, gene expression profiles generated by random sampling and sequencing is used for detailed genetic-level comparison between developmental, normal and pathological states of
10 the tissue examined.

Many human genes are expressed at different levels in cartilage of different developmental (fetal vs. mature) or disease states. In some cases, a gene is not expressed at all in some developmental or disease states, and at high levels in others (see Figure 6, 15 and 16 for examples). According to the invention, differential analysis of chondrocyte gene expression during different stages of cartilage developmental and in different disease states using an EST-based approach has identified genes that play important roles in osteoarthritis pathogenesis and cartilage repair. The advantage of this method is that it provides gene expression information on a larger scale than other methods. The cDNA clones generated by this approach are also useful for functional studies of certain genes. This type of genomic-
20 based approach has provided important novel insights into our understanding of the osteoarthritis disease process and provides for novel diagnostic, prognostic and therapeutic approaches.

Samples

Cartilage

25 In one aspect, cartilage is obtained from a fetus using methods known in the art. The chondrocytes of fetal cartilage have a higher level of metabolic activity and cell division rates

as compared to chondrocytes from cartilage from either a normal adult or from an individual diagnosed with any stage of osteoarthritis (mild, moderate, marked and severe).

In another aspect, cartilage is obtained from a normal individual who is alive or is obtained from cartilage tissue less than 14 hours post mortem, according to methods known in the art and described below. Normal articular cartilage from human adults are obtained using any known method. However, truly normal cartilage cannot generally be sampled from live donors due to ethical considerations. Preferably, normal cartilage samples are obtained from deceased donors, within a fourteen-hour post-mortem window, after cessation of perfusion to the sampled joint, minimize the degradation of RNA observed beyond the window. In other embodiments, the “normal” tissue is obtained less than 14 hours post-mortem, such as 13, 12, 11, 10, 9, 8, 6, 4, 2, or 1 hour post-mortem. A baboon study was conducted to confirm this approach and is described herein below in Example 11. Preferably the normal cartilage is obtained less than 14 hours post-mortem. More preferably, the normal cartilage is obtained less than 12 hours post-mortem.

Preferably, cartilage also is isolated from the following disease stages of osteoarthritis: mild, marked, moderate and severe. Human cartilage samples from osteoarthritic individuals are obtained using any known method. Preferably the cartilage is obtained from individuals undergoing arthroscopy or total knee replacements and samples are stored in liquid nitrogen until needed. In a preferred embodiment, a minimum of 0.05 g of cartilage sample is isolated to obtain 2 μ g total RNA extract for the construction of a cDNA library. In another preferred embodiment, a minimum of 0.025 g cartilage sample is isolated to obtain 1 μ g total RNA extract to use as a target sample for a microarray. A cartilage sample that is useful according to the invention is in an amount that is sufficient for the detection of one or more polynucleotide sequences according to the invention.

25

Blood and Synovial Fluid

Samples useful according to the invention also include blood and synovial fluid samples.

In one aspect, blood is obtained from a normal patient or from an individual diagnosed with, or suspected of having, osteoarthritis according to methods of phlebotomy well known in the art. A blood sample useful according to the invention is in an amount ranging from 1 μ l to 100ml, preferably 10 μ l to 50 ml, more preferably 10 μ l to 25ml and most preferably 10 μ l to 1 ml. A blood sample that is useful according to the invention is in an amount that is sufficient for the detection of one or more nucleic acid sequences according to the invention. In one embodiment, nucleic acids contained within the blood sample are amplified, for example, by polymerase chain reaction (PCR) or by RT-PCR. Other amplification methods known in the art are also encompassed within the scope of the invention (e.g., ligase chain reaction, NASBA, 3SR, and the like).

A synovial fluid sample is obtained from an individual diagnosed with, or suspected of having osteoarthritis according to methods well known in the art. Preferably, synovial fluid is collected from a human knee joint by aspiration at arthroscopy. A synovial fluid sample useful according to the invention is in an amount ranging from 0.1 ml to 20 ml and preferably 0.5 ml to 10 ml. A synovial fluid sample that is useful according to the invention is in an amount that is sufficient for the detection of one or more nucleic acid sequences according to the invention.

Developmental and Disease Stages of Articular Cartilage

Chondrocytes are preferably obtained from any of the following developmental and disease stages: fetal, normal, mild osteoarthritic, moderate osteoarthritic, marked osteoarthritic or severe osteoarthritic.

Cartilage isolated from a human fetus (e.g., during fetal development) is characterized above, and is useful according to the invention for analysis of fetal chondrocytes.

Cartilage isolated from a “normal” individual, defined herein, also is useful according to the invention for isolation and analysis of “normal” chondrocytes.

Cartilage isolated from a patient diagnosed with any one of: mild, moderate, marked and severe osteoarthritis also is useful in the present invention.

In order to classify cartilage according to disease state, a scoring system is used, whereby subjective decisions by the arthroscopist are minimized. The scoring system which 5 defines disease states described herein is that of Marshall, *supra*, incorporated herein by reference. According to this method, each of the 6 articular surfaces (patella, femoral trochlea, medial femoral condyle, medial tibial plateau, lateral femoral condyle and lateral tibial plateau) is assigned a cartilage grade based on the worst lesion present on that specific surface. A scoring system is then applied in which each articular surface receives an 10 osteoarthritis severity number value that reflects the cartilage severity grade for that surface, as described in Table 1.

Table 1. Articular Cartilage Grading System

Grade	Articular Cartilage	Points
0	Normal	0
I	Surface intact-softening, edema	1
II	Surface-disrupted-partial thickness lesions (no extension to bone)	2
III	Full thickness lesions-extensions to intact bone	3
IV	Bone erosion or eburnation	4

For example, if the medial femoral condyle has a grade I lesion as its most severe cartilage damage a value of 1 is assigned. A total score for the patient is then derived from the sum of the scores of the 6 articular surfaces. Based on the total score, each patient is 15 placed into one of 4 osteoarthritis groups: mild (1-6), moderate (7-12), marked (13-18) and severe (>18).

RNA Preparation

In one aspect, RNA is isolated from cartilage samples from various disease or developmental stages as described herein. Samples can be from single patients or can be pooled from multiple patients.

5 In another aspect, RNA is isolated directly from synovial fluid of persons with various disease or developmental stages of osteoarthritis as described herein. Samples can be from single patients or can be pooled from multiple patients.

In another aspect, RNA is isolated directly from blood samples of persons with various disease or developmental stages of osteoarthritis as described herein. Samples can be from single patients or can be pooled from multiple patients.

10 Total RNA is extracted from the cartilage samples according to methods well known in the art. In one embodiment, RNA is purified from cartilage tissue according to the following method. Following removal of a tissue of interest from an individual or patient, the tissue is quick frozen in liquid nitrogen, to prevent degradation of RNA. Upon the addition of a volume of tissue guanidinium solution, tissue samples are ground in a tissuemizer with two or three 10-second bursts. To prepare tissue guanidinium solution (1 L) 590.8 g guanidinium isothiocyanate is dissolved in approximately 400 ml DEPC-treated H₂O. 25 ml of 2 M Tris-Cl, pH 7.5 (0.05 M final) and 20 ml Na₂EDTA (0.01 M final) is added, the solution is stirred overnight, the volume is adjusted to 950 ml, and 50 ml 2-ME is added.

20 Homogenized tissue samples are subjected to centrifugation for 10 min at 12,000 x g at 12°C. The resulting supernatant is incubated for 2 min at 65°C in the presence of 0.1 volume of 20% Sarkosyl, layered over 9 ml of a 5.7M CsCl solution (0.1g CsCl/ml), and separated by centrifugation overnight at 113,000 x g at 22°C. After careful removal of the supernatant, the tube is inverted and drained. The bottom of the tube (containing the RNA pellet) is placed in a 50 ml plastic tube and incubated overnight (or longer) at 4°C in the presence of 3 ml tissue resuspension buffer (5 mM EDTA, 0.5% (v/v) Sarkosyl, 5% (v/v) 2-ME) to allow complete resuspension of the RNA pellet. The resulting RNA solution is extracted sequentially with 25:24:1 phenol/chloroform/isoamyl alcohol, followed by 24:1

chloroform/isoamyl alcohol, precipitated by the addition of 3 M sodium acetate, pH 5.2, and 2.5 volumes of 100% ethanol, and resuspended in DEPC water (Chirgwin et al., 1979, *Biochemistry*, 18:5294).

Alternatively, RNA is isolated from cartilage tissue according to the following single
5 step protocol. The tissue of interest is prepared by homogenization in a glass teflon
homogenizer in 1 ml denaturing solution (4M guanidinium thiosulfate, 25 mM sodium
citrate, pH 7.0, 0.1M 2-ME, 0.5% (w/v) N-laurylsarkosine) per 100mg tissue. Following
transfer of the homogenate to a 5-ml polypropylene tube, 0.1 ml of 2 M sodium acetate, pH 4,
10 1 ml water-saturated phenol, and 0.2 ml of 49:1 chloroform/isoamyl alcohol are added
sequentially. The sample is mixed after the addition of each component, and incubated for 15
min at 0-4°C after all components have been added. The sample is separated by
centrifugation for 20 min at 10,000 x g, 4°C, precipitated by the addition of 1 ml of 100%
15 isopropanol, incubated for 30 minutes at -20°C and pelleted by centrifugation for 10 minutes
at 10,000 x g, 4°C. The resulting RNA pellet is dissolved in 0.3 ml denaturing solution,
transferred to a microfuge tube, precipitated by the addition of 0.3 ml of 100% isopropanol
for 30 minutes at -20°C, and centrifuged for 10 minutes at 10,000 x g at 4°C. The RNA
pellet is washed in 70% ethanol, dried, and resuspended in 100-200 μ l DEPC-treated water or
DEPC-treated 0.5% SDS (Chomczynski and Sacchi, 1987, *Anal. Biochem.*, 162:156).

20 Preferably, the cartilage samples are finely powdered under liquid nitrogen and total
RNA is extracted using TRIzol® reagent (GIBCO/BRL).

Alternatively, RNA is isolated from blood by the following protocol. Lysis Buffer is
added to blood sample in a ratio of 3 parts Lysis Buffer to 1 part blood (Lysis Buffer (1L)
0.6g EDTA; 1.0g KHCO₂, 8.2g NH₄Cl adjusted to pH 7.4 (using NaOH)). Sample is mixed
and placed on ice for 5-10 minutes until transparent. Lysed sample is centrifuged at 1000
25 rpm for 10 minutes at 4°C, and supernatant is aspirated. Pellet is resuspended in 5ml Lysis
Buffer, and centrifuged again at 1000 rpm for 10 minutes at 4°C. Pelleted cells are
homogenized using TRIzol® (GIBCO/BRL) in a ratio of approximately 6ml of TRIzol® for

every 10ml of the original blood sample and vortexed well. Samples are left for 5 minutes at room temperature. RNA is extracted using 1.2 ml of chloroform per 1 ml of TRIzol®. Sample is centrifuged at 12,000 x g for 5 minutes at 4°C and upper layer is collected. To upper layer, isopropanol is added in ratio of 0.5 ml per 1 ml of TRIzol®. Sample is left 5 overnight at -20°C or for one hour at -20°C. RNA is pelleted in accordance with known methods, RNA pellet air dried, and pellet resuspended in DEPC treated ddH₂O. RNA samples can also be stored in 75% ethanol where the samples are stable at room temperature for transportation.

10 Alternatively, RNA is isolated from synovial fluid using TRIzol® reagent (GIBCO/BRL).

15 Purity and integrity of RNA is assessed by absorbance at 260/280nm and agarose gel electrophoresis followed by inspection under ultraviolet light.

Construction of cDNA libraries

20 cDNA libraries are constructed according to methods well known in the art (see for example Ausubel, *supra*, and Sambrook, *supra*, incorporated herein by reference).

25 In one aspect, cDNA samples, i.e., DNA that is complementary to RNA such as mRNA are prepared. The preparation of cDNA is well-known and well-documented in the prior art.

30 cDNA may be prepared according to the following method. Total cellular RNA is isolated (as described) and passed through a column of oligo(dT)-cellulose to isolate polyA RNA. The bound polyA mRNAs are eluted from the column with a low ionic strength buffer. To produce cDNA molecules, short deoxythymidine oligonucleotides (12-20 nucleotides) are hybridized to the polyA tails to be used as primers for reverse transcriptase, an enzyme that uses RNA as a template for DNA synthesis. Alternatively, or additionally, mRNA species are primed from many positions by using short oligonucleotide fragments comprising numerous sequences complementary to the mRNA of interest as primers for cDNA synthesis. The

resultant RNA-DNA hybrid is converted to a double stranded DNA molecule by a variety of enzymatic steps well-known in the art (Watson et al., 1992, *Recombinant DNA*, 2nd edition, Scientific American Books, New York).

To construct a cDNA library, the poly (A)⁺ RNA fraction may be isolated by oligo-dT cellulose chromatography (Pharmacia), and 3-5 ug poly (A)⁺ RNA is used to construct a cDNA library in the λ ZAP Express vector (Stratagene). Alternatively, cDNA libraries may be constructed into λ TriplEx2 vector through a PCR-based method, using SMART (Switching Mechanism At 5' end of RNA Transcript) cDNA Library Construction Kit (Clontech). First-strand cDNA is synthesized with an Xho I-oligo (dT) adapter-primer in the presence of 5'-methyl dCTP. After second-strand synthesis and ligation of EcoRI adapters, the cDNAs are digested with Xho I, resulting in cDNA flanked by EcoRI sites at the 5'-ends and Xho I sites at the 3'-ends. Digested cDNAs are size-fractionated in Sephadryl S-500 spin columns (Stratagene), then ligated into the λ ZAP Express vector predigested with EcoRI and Xho I. The resulting DNA/cDNA concatomers are packaged using Gigapack Gold packaging extracts. After titration, aliquots of primary packaging mix are stored in 7% DMSO at -80°C as primary library stocks, and the rest are amplified to establish stable library stocks.

From the amplified library, phage plaques are plated onto an appropriate medium. Preferably, phage plaques are plated at a density of 200-500 pfu/150 mm plate onto an *Escherichia coli* XL1-blue MRF' lawn with IPTG/X-gal for color selection. The plaques are then randomly picked and positive inserts are identified by polymerase chain reaction (PCR), according to methods well known in the art and described hereinbelow. Preferably, plaques are picked into 75 ul suspension media buffer (100 mM NaCl, 10 mM MgSO₄, 1 mM Tris, pH7.5, 0.02% gelatin). Phage elutes (5 ul) may be used for PCR reactions (50 ul total volume) with 125 umol/L of each dNTP (Pharmacia), 10 pmol each of modified T3 (5'-GCCAAGCTCGAAATTAAACCCCTCACTAAAGGG-3') and T7 (5'-CCAGTGAATTGTAATACGACTCACTATAGGGCG-3') primers, and 2 U of Taq DNA polymerase (Pharmacia). Reactions are cycled in a DNA Thermal Cycler (Perkin-Elmer) [denaturation at 95°C for 5 minutes, followed by 30 cycles of amplification (94°C, 45

seconds; 55°C, 30 seconds; 72°C, 3 minutes) and a terminal isothermal extension (72°C, 3 minutes)]. Agarose gel electrophoresis is used to assess the presence and purity of inserts.

The PCR product is then subjected to DNA sequencing using known methods (see Ausubel et al., *supra* and Sambrook et al., *supra*). Methods of sequencing employ such enzymes as the Klenow fragment of DNA polymerase I, Sequenase® (US Biochemical Corp, Cleveland, OH), Taq polymerase (Perkin Elmer, Norwalk, CT), thermostable T7 polymerase (Amersham, Chicago, IL), or combinations of recombinant polymerases and proofreading exonucleases such as the ELONGASE Amplification System (Gibco BRL, Gaithersburg, MD). Preferably, the process is automated with machines such as the Hamilton Micro Lab 10 2200 (Hamilton, Reno NV), Peltier Thermal Cycler (PTC200; MJ Research, Watertown, MA), the ABI 377 DNA sequencers (Perkin Elmer), and the PE Biosystems ABI Prism 3700 DNA Analyzer..

PCR products are first subjected to DNA sequencing reactions using specific primers, BigDye™ Terminator Cycle Sequencing v2.0 Ready Reaction (PE Biosystems), Tris MgCl buffer and water in a thermocycler. Sequencing reactions were incubated at 94°C for 2 minutes, followed by 25 cycles of 94°C, 30 seconds; 55°C, 20 seconds; and 72°C, 1 minute; and 15 cycles of 94°C, 30 seconds; and 72°C for 1 minute; and 72°C for 5 minutes. Reactions were then put on hold at 4°C until purified using methods well known in the prior art (i.e. alcohol precipitation or ethanol precipitation). Automated sequencing is preferably carried out with a PE Biosystems ABI Prism 3700 DNA Analyzer.

PCR

In one aspect, nucleic acid sequences of the invention are amplified by the polymerase chain reaction (PCR). PCR methods are well-known to those skilled in the art.

PCR provides a method for rapidly amplifying a particular nucleic acid sequence by 25 using multiple cycles of DNA replication catalyzed by a thermostable, DNA-dependent DNA polymerase to amplify the target sequence of interest. PCR requires the presence of a nucleic

acid to be amplified, two single-stranded oligonucleotide primers flanking the sequence to be amplified, a DNA polymerase, deoxyribonucleoside triphosphates, a buffer and salts.

The method of PCR is well known in the art. PCR, is performed as described in Mullis and Faloona, 1987, *Methods Enzymol.*, 155: 335, herein incorporated by reference.

5 PCR is performed using template DNA (at least 1fg; more usefully, 1-1000 ng) and at least 25 pmol of oligonucleotide primers. A typical reaction mixture includes: 2 μ l of DNA, 25 pmol of oligonucleotide primer, 2.5 μ l of 10H PCR buffer 1 (Perkin-Elmer, Foster City, CA), 0.4 μ l of 1.25 μ M dNTP, 0.15 μ l (or 2.5 units) of Taq DNA polymerase (Perkin Elmer, Foster City, CA) and deionized water to a total volume of 25 μ l. Mineral oil is overlaid and the PCR is performed using a programmable thermal cycler.

10 The length and temperature of each step of a PCR cycle, as well as the number of cycles, are adjusted according to the stringency requirements in effect. Annealing temperature and timing are determined both by the efficiency with which a primer is expected to anneal to a template and the degree of mismatch that is to be tolerated. The ability to 15 optimize the stringency of primer annealing conditions is well within the knowledge of one of moderate skill in the art. An annealing temperature of between 30°C and 72°C is used. Initial denaturation of the template molecules normally occurs at between 92°C and 99°C for 4 minutes, followed by 20-40 cycles consisting of denaturation (94-99°C for 15 seconds to 1 minute), annealing (temperature determined as discussed above; 1-2 minutes), and extension 20 (72°C for 1 minute). The final extension step is generally carried out for 4 minutes at 72°C, and may be followed by an indefinite (0-24 hour) step at 4°C.

25 Several techniques for detecting PCR products quantitatively without electrophoresis may be useful according to the invention . One of these techniques, for which there are commercially available kits such as Taqman™ (Perkin Elmer, Foster City, CA), is performed with a transcript-specific antisense probe. This probe is specific for the PCR product (e.g. a nucleic acid fragment derived from a gene) and is prepared with a quencher and fluorescent reporter probe complexed to the 5' end of the oligonucleotide. Different fluorescent markers

are attached to different reporters, allowing for measurement of two products in one reaction. When Taq DNA polymerase is activated, it cleaves off the fluorescent reporters of the probe bound to the template by virtue of its 5'-to-3' exonuclease activity. In the absence of the quenchers, the reporters now fluoresce. The color change in the reporters is proportional to 5 the amount of each specific product and is measured by a fluorometer; therefore, the amount of each color is measured and the PCR product is quantified. The PCR reactions are performed in 96 well plates so that samples derived from many individuals are processed and measured simultaneously. The Taqman™ system has the additional advantage of not requiring gel electrophoresis and allows for quantification when used with a standard curve.

10 Nucleic acid Sequences Useful According to the Invention

The invention provides for isolated nucleic acid sequences including ESTs which can be used as probes, arrayed on microarrays, and/or used for the development of therapies to treat osteoarthritis.

15 In one aspect, cartilage gene expression profiles at different developmental stages are identified. Another aspect of the invention is to monitor cartilage gene expression profiles of osteoarthritis patients diagnosed with different stages of osteoarthritis. A third aspect of the invention is to screen for potential therapeutic agents which alter the gene expression profile of diseased cartilage cells. The invention therefore provides for nucleic acid sequences that are present at each of the following developmental and disease stages: normal, fetal, mild
20 osteoarthritic, moderate osteoarthritic, marked osteoarthritic and severe osteoarthritic. The invention also provides for nucleic acid sequences that are differentially expressed in any two of the following developmental and disease stages: normal, fetal, mild osteoarthritic, moderate osteoarthritic, marked osteoarthritic and severe osteoarthritic.

25 Nucleic acids useful according to the invention are prepared by isolating cartilage tissue samples from a developmental or disease stage (normal, fetal, mild osteoarthritic, moderate osteoarthritic, marked osteoarthritic and severe osteoarthritic), preparing a cDNA library (as described above), and performing large-scale partial sequencing (described herein)

of the cDNA library to generate Expressed Sequence Tags (ESTs). An EST useful according to the invention is preferably in the range of 50-1000 nucleotides and most preferably 100-500 nucleotides in length.

The invention provides for nucleic acid sequences or ESTs that are categorized as 5 "novel" or "known", including "known sequences with a function" and "known sequences without a known function", all defined herein.

Nucleic Acid Members and Probes

In one aspect, the invention provides nucleic acid members and probes that bind specifically to a target nucleic acid sequence (e.g., present in a cartilage nucleic acid sample).

10 Nucleic acid members are stably associated with a solid support to comprise an array according to the invention. The length of a nucleic acid member can range from 50 to 6000 nucleotides, 100 to 500 nucleotides, and in other embodiments, from 500 to 1500 nucleotides. The nucleic acid members may be single or double stranded, and/or may be PCR fragments amplified from cDNA.

15 The invention also provides for nucleic acid sequences comprising a probe. In a certain embodiment, a probe is labeled, according to methods known in the art. A probe according to the invention is 50 to 5000 nucleotides, more preferably 100-500 nucleotides and most preferably 50 to 250 nucleotides in length. The probe may be single or double stranded, and may be a PCR fragment amplified from cDNA.

20 The nucleic acid members and probes according to the invention can be used to detect target sequences such as chondrocyte enriched or chondrocyte-specific sequences, and preferably sequences whose presence in a sample are indicative, or diagnostic or prognostic, of a stage of osteoarthritis.

The target nucleic acid sequences to be analyzed are preferably from human cartilage, blood or synovial fluid and preferably comprise RNA or nucleic acid corresponding to RNA, (i.e., cDNA or amplified products of RNA or cDNAs).

Data Acquisition and Analysis of EST Sequences

5 The invention provides for EST sequences including “novel sequences”, “novel expressed sequence tags (ESTs)” and “known sequences” including “known sequences with a function” and “known sequences with no known function”.

10 The generated EST sequences are searched against available databases, including the “nt”, “nr”, “est”, “gss” and “htg” databases available through NCBI to determine putative identities for ESTs matching to known genes or other ESTs. Relative EST frequency level can then be calculated using known methods. Functional characterization of ESTs with known gene matches are made according to any known method. Preferably, generated EST sequences are compared to the non-redundant Genbank/EMBL/DDBJ and dbEST databases using the BLAST algorithm (8). A minimum value of $P = 10^{-10}$ and nucleotide sequence identity >95%, where the sequence identity is non-contiguous or scattered, are required for assignments of putative identities for ESTs matching to known genes or to other ESTs. Construction of a non-redundant list of genes represented in the EST set is done with the help of Unigene, Entrez and PubMed at the National Center for Biotechnology Information (NCBI) site (<http://www.ncbi.nlm.nih.gov/>). Relative gene expression frequency is calculated
15 by dividing the number of EST copies for each gene by the total number of ESTs analyzed.

20 Genes are identified from ESTs according to known methods. To identify novel genes from an EST sequence, the EST should preferably be at least 100 nucleotides in length, and more preferably 150 nucleotides in length, for annotation. Preferably, the EST exhibits open reading frame characteristics (i.e., can encode a putative polypeptide).

25 Because of the completion of the Human Genome Project, a specific EST which matches with a genomic sequence can be mapped onto a specific chromosome based on the chromosomal location of the genomic sequence. However, no function may be known for the

protein encoded by the sequence and the EST would then be considered "novel" in a functional sense. In one aspect, the invention is used to identify a novel EST which is part of a larger known sequence for which no function is known is used to determine the function of a gene comprising the EST (e.g., such as the role of expression products produced by the gene
5 in chondrogenesis and/or in a pathology affecting chondrocytes). Alternatively, or additionally, the EST can be used to identify an mRNA or polypeptide encoded by the larger sequence as a diagnostic or prognostic marker of chondrogenesis and/or of a pathology affecting chondrocytes.

Having identified an EST corresponding to a larger sequence as chondrocyte enriched or chondrocyte-specific, other portions of the larger sequence which comprises the EST can be used in assays to elucidate gene function, e.g., to isolate polypeptides encoded by the gene, to generate antibodies specifically reactive with these polypeptides, to identify binding partners of the polypeptides (receptors, ligands, agonists, antagonists and the like) and/or to detect the expression of the gene (or lack thereof) in chondrocytes in fetal, adult, normal, and/or diseased individuals.
10
15

In another aspect, the invention provides for nucleic acid sequences that do not demonstrate a "significant match" to any of the publicly known sequences in sequence databases at the time a query is done. Longer genomic segments comprising these types of novel EST sequences can be identified by probing genomic libraries, while longer expressed sequences can be identified in cDNA libraries and/or by performing polymerase extension reactions (e.g., RACE) using EST sequences to derive primer sequences as is known in the art. Longer fragments can be mapped to particular chromosomes by FISH and other techniques and their sequences compared to known sequences in genomic and/or expressed sequence databases and further functional analysis can be performed as described above.
20

Using the methods according to the invention, out of a total of 57,422 ESTs from the four cDNA libraries, no significant match was found for 618 sequences. The remaining sequences were characterized as shown in Figures 5 and 5A.
25

Identified genes can be catalogued according to their putative function. Functional characterization of ESTs with known gene matches is preferably made according to the categories described by Hwang et al (5). The distribution of genes in each of the subcellular categories is indicative of the dynamic state of the tissue and will provide important insights 5 into the osteoarthritis disease process. The results of this analysis are provided in Figures 7 and 7A where the total number of ESTs identified by the method in different human cartilage libraries are characterized based on the functional classification of known genes identified in each library.

Alternative methods for analyzing ESTs are also available. For example, the ESTs 10 from each library may be assembled into contigs with sequence alignment, editing, and assembly programs such as PHRED and PHRAP (Ewing, et al., 1998, *Genome Res.* 3:175, incorporated herein; <http://bozeman.genome.washington.edu/>). Contig redundancy is reduced by clustering nonoverlapping sequence contigs using the EST clone identification number, which is common for the nonoverlapping 5 and 3 sequence reads for a single EST cDNA 15 clone. In one aspect, the consensus sequence from each cluster is compared to the non-redundant Genbank/EMBL/DDBJ and dbEST databases using the BLAST algorithm with the help of unigene, Entrez and PubMed at the NCBI site.

Known Nucleic acid Sequences or ESTs and Novel Nucleic acid Sequences or ESTs

An EST that exhibits a significant match (> 65%, and preferably 90% or greater, 20 identity) to at least one existing sequence in an existing nucleic acid sequence database is characterized as a “known” sequence according to the invention. Within this category, some known ESTs match to existing sequences which encode polypeptides with known function(s) and are referred to as a “known sequence with a function”. Other “known” ESTs exhibit significant match to existing sequences which encode polypeptides of unknown function(s) 25 and are referred to as a “known sequence with no known function”.

In one aspect, the invention also provides for known nucleic acid sequences that are chondrocyte enriched or chondrocyte-specific.

EST sequences which have no significant match (less than 65% identity) to any existing sequence in the above cited available databases are categorized as novel ESTs. These novel ESTs are considered chondrocyte-specific since they are not matched to any other genes or ESTs derived from any other tissue. To identify a novel gene from an EST 5 sequence, the EST is preferably at least 150 nucleotides in length. More preferably, the EST encodes at least part of an open reading frame, that is, a nucleic acid sequence between a translation initiation codon and a termination codon, which is potentially translated into a polypeptide sequence.

10 The invention provides for known and novel nucleic acid sequences that are uniquely expressed in normal, fetal, mild osteoarthritic, moderate osteoarthritic, marked osteoarthritic and severe osteoarthritic cartilage. Figures 6 and 13, show unique known genes and names of the novel sequences identified to date in the fetal, normal, mild osteoarthritic and severe 15 osteoarthritic cDNA libraries using the methods according to the invention.

15 The invention also provides for known and novel nucleic acid sequences that are upregulated and downregulated in normal, fetal, mild osteoarthritic, moderate osteoarthritic, marked osteoarthritic and severe osteoarthritic cartilage. In one aspect, nucleic acid sequences are enriched in chondrocytes compared to cells which are non-chondrocytes, or in 20 chondrocytes from individuals with osteoarthritis compared to normal individuals, or in chondrocytes from particular stages of development or disease compared to particular other stages of development or disease.

The invention also provides for nucleic acid sequences that are differentially expressed in cartilage from any two of the following developmental and disease stages: fetal, normal, mild osteoarthritic, moderate osteoarthritic, marked osteoarthritic and severe osteoarthritic.

25 Relative EST frequency is calculated by dividing the number of EST copies for each gene by the total number of ESTs analyzed. The chondrocyte-specific expression of a

number of novel ESTs has been confirmed by methods known in the art. Useful methods for measuring gene expression in a tissue include RT PCR, Northern blot, etc.

Novel Nucleic Acid Molecules

Many of the novel nucleic acid molecules of the present invention are differentially expressed between the mild and severe osteoarthritis disease states and are thus useful as potential drug targets or markers for the osteoarthritis disease process. The invention also provides one or more nucleic acid molecules that are differentially expressed in two or more of the following developmental and disease stages: fetal, normal, mild osteoarthritic, moderate osteoarthritic, marked osteoarthritic and severe osteoarthritic. The invention further provides one or more novel clones that are differentially expressed in two or more of the following developmental and disease stages: fetal, normal, mild osteoarthritic, moderate osteoarthritic, marked osteoarthritic and severe osteoarthritic.

Microarrays

Nucleic acid Microarrays

Any combination of the nucleic acid sequences generated from any of the chondrocyte cDNA libraries are used for the construction of a microarray. In one embodiment, the microarray is chondrocyte-specific and is anticipated to encompass the entire spectrum of genes that are important in the osteoarthritis disease process. A microarray according to the invention preferably comprises between 10 and 20,000 nucleic acid members, and more preferably comprises at least 5000 nucleic acid members. The nucleic acid members are known or novel nucleic acid sequences described herein, or any combination thereof. A microarray according to the invention is used to confirm differential gene expression profiles of genes that are specifically expressed at different cartilage development and osteoarthritis disease stages.

The invention also provides for a microarray comprising genes that are differentially expressed between normal and mild osteoarthritis patients to allow for the identification of

early risk factors for osteoarthritis development. The invention also provides for a microarray for osteoarthritis diagnosis comprising one or more nucleic acid sequences that are differentially expressed between a normal individual and a patient diagnosed with mild, moderate, marked or severe osteoarthritis. Such arrays also may be used for prognostic 5 methods to monitor a patient's response to therapy. Preferably, an array for osteoarthritis diagnosis comprises 10-20,000 nucleic acid members and more preferably 50-15,000 nucleic acid members. In one embodiment, the above microarrays are used to identify a therapeutic agent that modulates the anabolic activity of a chondrocyte or changes (e.g., increases or decreases) the level of expression of at least one nucleic acid sequence that is differentially 10 expressed in a chondrocyte derived from any of the following chondrocyte disease or developmental stages: fetal, normal, mild osteoarthritic, moderate osteoarthritic, marked osteoarthritic and severe osteoarthritic.

15 The target nucleic acid samples that are hybridized to and analyzed with a microarray of the invention are preferably from human cartilage, blood or synovial fluid. A limitation for this procedure lies in the amount of RNA available for use as a target nucleic acid sample. Preferably, at least 1 microgram of total RNA is obtained for use according to this invention. This is advantageous because the amount of RNA in synovial fluid and in many cartilage biopsy samples is very minimal.

Construction of a Microarray

20 In one aspect, cDNAs generated from human cartilage cDNA libraries are arrayed on a microarray. Preferably, a microarray according to the invention comprises chondrocyte enriched or chondrocyte-specific genes and includes the whole spectrum of genes that are important in the osteoarthritis disease process.

25 The EST frequency analysis in Figure 6 (and portions thereof shown in Figures 15 and 16) shows the differential gene expression profiles for known genes. Microarrays according to the invention may be used to confirm these profiles and may also be used to show differential expression profiles between different developmental stages and osteoarthritis

disease states for novel EST sequences. These novel EST sequences may be further characterized by cluster and alignment analyses to determine how many unique genes are represented by the novel EST sequences. The novel unique genes identified may provide a basis for identifying key markers in osteoarthritis disease progression and treatment.

5 In the subject methods, an array of nucleic acid members stably associated with the surface of a substantially solid support is contacted with a sample comprising target nucleic acids under hybridization conditions sufficient to produce a hybridization pattern of complementary nucleic acid members/target complexes in which one or more complementary nucleic acid members at unique positions on the array specifically hybridize to target nucleic
10 acids. The identity of target nucleic acids which hybridize can be determined with reference to location of nucleic acid members on the array.

15 The nucleic acid members may be produced using established techniques such as polymerase chain reaction (PCR) and reverse transcription (RT). These methods are similar to those currently known in the art (see e.g., *PCR Strategies*, Michael A. Innis (Editor), et al. (1995) and *PCR: Introduction to Biotechniques Series*, C. R. Newton, A. Graham (1997)). Amplified nucleic acids are purified by methods well known in the art (e.g., column purification or alcohol precipitation). A nucleic acid is considered pure when it has been isolated so as to be substantially free of primers and incomplete products produced during the synthesis of the desired nucleic acid. Preferably, a purified nucleic acid will also be
20 substantially free of contaminants which may hinder or otherwise mask the specific binding activity of the molecule.

25 A microarray according to the invention comprises a plurality of unique nucleic acids attached to one surface of a solid support at a density exceeding 20 different nucleic acids/cm², where each of the nucleic acids is attached to the surface of the solid support in a non-identical pre-selected region. Each associated sample on the array comprises a nucleic acid composition, of known identity, usually of known sequence, as described in greater detail below. Any conceivable substrate may be employed in the invention.

In one embodiment, the nucleic acid attached to the surface of the solid support is DNA. In a preferred embodiment, the nucleic acid attached to the surface of the solid support is cDNA or RNA. In another preferred embodiment, the nucleic acid attached to the surface of the solid support is cDNA synthesized by polymerase chain reaction (PCR). Preferably, a 5 nucleic acid member in the array, according to the invention, is at least 50 nucleotides in length. In one embodiment, a nucleic acid member is at least 150 nucleotides in length. Preferably, a nucleic acid member is less than 1000 nucleotides in length. More preferably, a nucleic acid member is less than 500 nucleotides in length. In one embodiment, an array 10 comprises at least 10 different nucleic acids attached to one surface of the solid support. In another embodiment, the array comprises at least 100 different nucleic acids attached to one surface of the solid support. In yet another embodiment, the array comprises at least 10,000 different nucleic acids attached to one surface of the solid support. In yet another embodiment, the array comprises at least 15,000 different nucleic acids attached to one surface of the solid support.

15 In the arrays of the invention, the nucleic acid compositions are stably associated with the surface of a solid support, where the support may be a flexible or rigid solid support. By “stably associated” is meant that each nucleic acid member maintains a unique position relative to the solid support under hybridization and washing conditions. As such, the samples are non-covalently or covalently stably associated with the support surface. 20 Examples of non-covalent association include non-specific adsorption, binding based on electrostatic interactions (e.g., ion pair interactions), hydrophobic interactions, hydrogen bonding interactions, specific binding through a specific binding pair member covalently attached to the support surface, and the like. Examples of covalent binding include covalent bonds formed between the nucleic acids and a functional group present on the surface of the 25 rigid support (e.g., --OH), where the functional group may be naturally occurring or present as a member of an introduced linking group, as described in greater detail below

The amount of nucleic acid present in each composition will be sufficient to provide for adequate hybridization and detection of target nucleic acid sequences during the assay in

which the array is employed. Generally, the amount of each nucleic acid member stably associated with the solid support of the array is at least about 0.001 ng, preferably at least about 0.02 ng and more preferably at least about 0.05 ng, where the amount may be as high as 1000 ng or higher, but will usually not exceed about 20 ng. Where the nucleic acid member 5 is “spotted” onto the solid support in a spot comprising an overall circular dimension, the diameter of the “spot” will generally range from about 10 to 5,000 μm , usually from about 20 to 2,000 μm and more usually from about 100 to 200 μm .

Control nucleic acid members may be present on the array including nucleic acid members comprising oligonucleotides or nucleic acids corresponding to genomic DNA, 10 housekeeping genes, vector sequences, plant nucleic acid sequence, negative and positive control genes, and the like. Control nucleic acid members are calibrating or control genes whose function is not to tell whether a particular “key” gene of interest is expressed, but rather to provide other useful information, such as background or basal level of expression.

Other control nucleic acids are spotted on the array and used as target expression 15 control nucleic acids and mismatch control nucleotides to monitor non-specific binding or cross-hybridization to a nucleic acid in the sample other than the target to which the probe is directed. Mismatch probes thus indicate whether a hybridization is specific or not. For example, if the target is present, the perfectly matched probes should be consistently brighter than the mismatched probes. In addition, if all control mismatches are present, the mismatch 20 probes are used to detect a mutation.

Solid Substrate

An array according to the invention comprises either a flexible or rigid substrate. A flexible substrate is capable of being bent, folded or similarly manipulated without breakage. Examples of solid materials which are flexible solid supports with respect to the present 25 invention include membranes, e.g., nylon, flexible plastic films, and the like. By “rigid” is meant that the support is solid and does not readily bend, i.e., the support is not flexible. As such, the rigid substrates of the subject arrays are sufficient to provide physical support and

structure to the associated nucleic acids present thereon under the assay conditions in which the array is employed, particularly under high throughput handling conditions.

The substrate may be biological, non-biological, organic, inorganic, or a combination of any of these, existing as particles, strands, precipitates, gels, sheets, tubing, spheres, beads, 5 containers, capillaries, pads, slices, films, plates, slides, chips, etc. The substrate may have any convenient shape, such as a disc, square, sphere, circle, etc. The substrate is preferably flat or planar but may take on a variety of alternative surface configurations. The substrate may be a polymerized Langmuir Blodgett film, functionalized glass, Si, Ge, GaAs, GaP, SiO₂, SiN₄, modified silicon, or any one of a wide variety of gels or polymers such as 10 (poly)tetrafluoroethylene, (poly)vinylidenedifluoride, polystyrene, polycarbonate, or combinations thereof. Other substrate materials will be readily apparent to those of skill in the art upon review of this disclosure.

In a preferred embodiment the substrate is flat glass or single-crystal silicon. According to some embodiments, the surface of the substrate is etched using well-known 15 techniques to provide for desired surface features. For example, by way of formation of trenches, v-grooves, mesa structures, or the like, the synthesis regions may be more closely placed within the focus point of impinging light, be provided with reflective "mirror" structures for maximization of light collection from fluorescent sources, etc.

Surfaces on the solid substrate will usually, though not always, be composed of the 20 same material as the substrate. Alternatively, the surface may be composed of any of a wide variety of materials, for example, polymers, plastics, resins, polysaccharides, silica or silica-based materials, carbon, metals, inorganic glasses, membranes, or any of the above-listed substrate materials. In some embodiments the surface may provide for the use of caged binding members which are attached firmly to the surface of the substrate. Preferably, the 25 surface will contain reactive groups, which are carboxyl, amino, hydroxyl, or the like. Most preferably, the surface will be optically transparent and will have surface Si--OH functionalities, such as are found on silica surfaces.

The surface of the substrate is preferably provided with a layer of linker molecules, although it will be understood that the linker molecules are not required elements of the invention. The linker molecules are preferably of sufficient length to permit nucleic acids of the invention and on a substrate to hybridize to other nucleic acid molecules and to interact 5 freely with molecules exposed to the substrate.

Often, the substrate is a silicon or glass surface, (poly)tetrafluoroethylene, (poly)vinylidendifluoride, polystyrene, polycarbonate, a charged membrane, such as nylon 66 or nitrocellulose, or combinations thereof. In a preferred embodiment, the solid support is glass. Preferably, at least one surface of the substrate will be substantially flat. Preferably, 10 the surface of the solid support will contain reactive groups, including, but not limited to, carboxyl, amino, hydroxyl, thiol, or the like. In one embodiment, the surface is optically transparent. In a preferred embodiment, the substrate is a poly-lysine coated slide or Gamma amino propyl silane-coated Corning Microarray Technology-GAPS or CMT-GAP2 coated slides.

15 Any solid support to which a nucleic acid member may be attached may be used in the invention. Examples of suitable solid support materials include, but are not limited to, silicates such as glass and silica gel, cellulose and nitrocellulose papers, nylon, polystyrene, polymethacrylate, latex, rubber, and fluorocarbon resins such as TEFLON™.

The solid support material may be used in a wide variety of shapes including, but not 20 limited to slides and beads. Slides provide several functional advantages and thus are a preferred form of solid support. Due to their flat surface, probe and hybridization reagents are minimized using glass slides. Slides also enable the targeted application of reagents, are easy to keep at a constant temperature, are easy to wash and facilitate the direct visualization of RNA and/or DNA immobilized on the solid support. Removal of RNA and/or DNA 25 immobilized on the solid support is also facilitated using slides.

The particular material selected as the solid support is not essential to the invention, as long as it provides the described function. Normally, those who make or use the invention

will select the best commercially available material based upon the economics of cost and availability, the expected application requirements of the final product, and the demands of the overall manufacturing process.

Spotting Method

5 In one aspect, The invention provides for arrays where each nucleic acid member comprising the array is spotted onto a solid support.

Preferably, spotting is carried out as follows. PCR products (~40 ul) of cDNA clones from osteoarthritis, fetal or normal cartilage cDNA libraries, in the same 96-well tubes used for amplification, are precipitated with 4 ul (1/10 volume) of 3M sodium acetate (pH 5.2) and 100 ul (2.5 volumes) of ethanol and stored overnight at -20°C. They are then centrifuged at 3,300 rpm at 4°C for 1 hour. The obtained pellets are washed with 50 ul ice-cold 70% ethanol and centrifuged again for 30 minutes. The pellets are then air-dried and resuspended well in 20ul 3X SSC or in 50% dimethylsulfoxide (DMSO) overnight. The samples are then spotted, either singly or in duplicate, onto polylysine-coated slides (Sigma Cat. No. P0425) using a robotic GMS 417 or 427 arrayer (Affymetrix, Ca).

The boundaries of the spots on the microarray may be marked with a diamond scriber (as the spots become invisible after post-processing). The arrays are rehydrated by suspending the slides over a dish of warm particle free ddH₂O for approximately one minute (the spots will swell slightly but will not run into each other) and snap-dried on a 70-80°C inverted heating block for 3 seconds. Nucleic acid is then UV crosslinked to the slide (Stratagene, Stratalinker, 65 mJ – set display to “650” which is 650 x 100 uJ) or the array is baked at 80C for two to four hours prior to hybridization. The arrays are placed in a slide rack. An empty slide chamber is prepared and filled with the following solution: 3.0 grams of succinic anhydride (Aldrich) was dissolved in 189 ml of 1-methyl-2-pyrrolidinone (rapid addition of reagent is crucial); immediately after the last flake of succinic anhydride is dissolved, -21.0 ml of 0.2 M sodium borate is mixed in and the solution is poured into the slide chamber. The slide rack is plunged rapidly and evenly in the slide chamber and

vigorously shaken up and down for a few seconds, making sure the slides never leave the solution, and then mixed on an orbital shaker for 15-20 minutes. The slide rack is then gently plunged in 95°C ddH₂O for 2 minutes, followed by plunging five times in 95% ethanol. The slides are then air dried by allowing excess ethanol to drip onto paper towels. The arrays are
5 stored in the slide box at room temperature until use.

Numerous methods may be used for attachment of the nucleic acid members of the invention to the substrate (a process referred to as "spotting"). For example, nucleic acids are attached using the techniques of, for example U.S. Pat. No. 5,807,522, which is incorporated herein by reference, for teaching methods of polymer attachment.

10 Alternatively, spotting may be carried out using contact printing technology as is known in the art.

Kits

The invention provides for kits for performing expression assays using the arrays of the present invention. Such kits according to the subject invention will at least comprise the arrays of the invention having associated nucleic acid members and packaging means therefore. The kits may further comprise one or more additional reagents employed in the various methods, such as: 1) primers for generating test nucleic acids; 2) dNTPs and/or rNTPs (either premixed or separate), optionally with one or more uniquely labeled dNTPs and/or rNTPs (e.g., biotinylated or Cy3 or Cy5 tagged dNTPs); 3) post synthesis labeling reagents, 15 such as chemically active derivatives of fluorescent dyes; 4) enzymes, such as reverse transcriptases, DNA polymerases, and the like; 5) various buffer mediums, e.g., hybridization and washing buffers; 6) labeled probe purification reagents and components, like spin columns, etc.; and 7) signal generation and detection reagents, e.g., streptavidin-alkaline phosphatase conjugate, chemiluminescent or chemiluminescent substrate, and the like.

20
25 Use of a Microarray

Nucleic acid arrays according to the invention can be used in high throughput techniques that can assay a large number of nucleic acids in a sample comprising one or more target nucleic acid sequences. The arrays of the subject invention find use in a variety of applications, including gene expression analysis, diagnosis of osteoarthritis and prognosis of 5 osteoarthritis (e.g., monitoring a patient's response to therapy, drug screening, and the like).

In one aspect, the arrays of the invention are used in, among other applications, differential gene expression assays. For example, arrays are useful in the differential expression analysis of: (a) diseased osteoarthritis and normal tissue; (b) tissues representing different stages of osteoarthritis; (c) developing cartilage (e.g., fetal cartilage); (d) 10 chondrocyte responses to external or internal stimuli; (e) cartilage/chondrocyte response to treatment; (f) cartilage tissue engineering; (g) pharmacogenomics; and the like. The arrays are also useful in broad scale expression screening for drug discovery and research, such as the effect of a particular active agent on the expression pattern of genes in a particular cell, where such information is used to reveal drug efficacy and toxicity, environmental monitoring, disease research and the like. For example, high expression of a particular 15 nucleic acid sequence in an osteoarthritis sample (mild, moderate, marked, or severe), which is not observed in a corresponding normal cell, can indicate an osteoarthritis-specific gene product.

Target Preparation

20 The targets for the microarrays according to the invention are preferably derived from human cartilage, blood or synovial fluid.

A target nucleic acid is capable of binding to a nucleic acid probe or nucleic acid member of complementary sequence through one or more types of chemical bonds, usually through complementary base pairing, usually through hydrogen bond formation.

25 As used herein, a "nucleic acid derived from an mRNA transcript: or a "nucleic acid corresponding to an mRNA" refers to a nucleic acid for which synthesis of the mRNA transcript or a sub-sequence thereof has ultimately served as a template. Thus, a cDNA

reverse transcribed from an mRNA, an RNA transcribed from that cDNA, a DNA amplified from the cDNA, an RNA transcribed from the amplified DNA, etc., are all derived from or correspond to the mRNA transcript and detection of such derived or corresponding products is indicative of or proportional to the presence and/or abundance of the original transcript in a sample. Thus, suitable target nucleic acid samples include, but are not limited to, mRNA transcripts of a gene or genes, cDNA reverse transcribed from the mRNA, cRNA transcribed from the cDNA, DNA amplified from a gene or genes, RNA transcribed from amplified DNA, and the like. The nucleic acid targets used herein are preferably derived from human cartilage, blood or synovial fluid. Preferably, the targets are nucleic acids derived from human cartilage, blood or synovial fluid extracts. Nucleic acids can be single- or double-stranded DNA, RNA, or DNA-RNA hybrids synthesized from human cartilage, blood or synovial fluid mRNA extracts using methods known in the art, for example, reverse transcription or PCR.

In the simplest embodiment, such a nucleic acid target comprises total mRNA or a nucleic acid sample corresponding to mRNA (e.g., cDNA) isolated from cartilage, blood, or synovial fluid samples. In another embodiment, total mRNA is isolated from a given sample using, for example, an acid guanidinium-phenol-chloroform extraction method and polyA+ mRNA is isolated by oligo dT column chromatography or by using (dT)_n magnetic beads (see, e.g., Sambrook et al., *Molecular Cloning: A Laboratory Manual* (2nd ed.), Vols. 1-3, Cold Spring Harbor Laboratory, (1989), or Current Protocols in Molecular Biology, F. Ausubel et al., ed. Greene Publishing and Wiley-Interscience, New York (1987). In a preferred embodiment, total RNA is extracted using TRIzol® reagent (GIBCO/BRL, Invitrogen Life Technologies, Cat. No. 15596). Purity and integrity of RNA is assessed by absorbance at 260/280nm and agarose gel electrophoresis followed by inspection under ultraviolet light.

In some embodiments, it is desirable to amplify the target nucleic acid sample prior to hybridization, for example, when synovial fluid is used. One of skill in the art will appreciate that whatever amplification method is used, if a quantitative result is desired, care must be

taken to use a method that maintains or controls for the relative frequencies of the amplified nucleic acids. Methods of "quantitative" amplification are well known to those of skill in the art. For example, quantitative PCR involves simultaneously co-amplifying a known quantity of a control sequence using the same primers. This provides an internal standard that may be 5 used to calibrate the PCR reaction. The high density array may then include probes specific to the internal standard for quantification of the amplified nucleic acid. Detailed protocols for quantitative PCR are provided in *PCR Protocols, A Guide to Methods and Applications*, Innis et al., Academic Press, Inc. N.Y., (1990).

Other suitable amplification methods include, but are not limited to polymerase chain 10 reaction (PCR) (Innis, et al., *PCR Protocols. A Guide to Methods and Application*. Academic Press, Inc. San Diego, (1990)), ligase chain reaction (LCR) (see Wu and Wallace, 1989, *Genomics*, 4:560; Landegren, et al., 1988, *Science*, 241:1077 and Barringer, et al., 1990, *Gene*, 89:117, transcription amplification (Kwoh, et al., 1989, *Proc. Natl. Acad. Sci. USA*, 86: 1173), and self-sustained sequence replication (Guatelli, et al., 1990, *Proc. Nat. Acad. Sci. USA*, 87: 1874).

In a particularly preferred embodiment, the target nucleic acid sample mRNA is reverse transcribed with a reverse transcriptase and a primer consisting of oligo dT and a sequence encoding the phage T7 promoter to provide single-stranded DNA template. The second DNA strand is polymerized using a DNA polymerase. After synthesis of double-stranded cDNA, T7 RNA polymerase is added and RNA is transcribed from the cDNA template. Successive rounds of transcription from each single cDNA template results in amplified RNA. Methods of *in vitro* transcription are well known to those of skill in the art (see, e.g., Sambrook, *supra*.) and this particular method is described in detail by Van Gelder, et al., 1990, *Proc. Natl. Acad. Sci. USA*, 87: 1663-1667 who demonstrate that *in vitro* 20 amplification according to this method preserves the relative frequencies of the various RNA transcripts. Moreover, Eberwine et al. *Proc. Natl. Acad. Sci. USA*, 89: 3010-3014 provide a protocol that uses two rounds of amplification via *in vitro* transcription to achieve greater 25

than 10^6 fold amplification of the original starting material thereby permitting expression monitoring even where biological samples are limited.

Labeling of Target or Nucleic Acid Probe

Either the target or the probe can be labeled.

5 Any analytically detectable marker that is attached to or incorporated into a molecule may be used in the invention. An analytically detectable marker refers to any molecule, moiety or atom which is analytically detected and quantified.

10 Detectable labels suitable for use in the present invention include any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical or chemical means. Useful labels in the present invention include biotin for staining with labeled streptavidin conjugate, magnetic beads (e.g., DynabeadsTM), fluorescent dyes (e.g., fluorescein, texas red, rhodamine, green fluorescent protein, and the like), radiolabels (e.g., ³H, ¹²⁵I, ³⁵S, ¹⁴C, or ³²P), enzymes (e.g., horse radish peroxidase, alkaline phosphatase and others commonly used in an ELISA), and colorimetric labels such as colloidal gold or colored glass or plastic (e.g., polystyrene, polypropylene, latex, etc.) beads. Patents teaching the use of such labels include U.S. Pat. Nos. 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 15 4,275,149; and 4,366,241, the entireties of which are incorporated by reference herein.

20 Means of detecting such labels are well known to those of skill in the art. Thus, for example, radiolabels may be detected using photographic film or scintillation counters, fluorescent markers may be detected using a photodetector to detect emitted light. Enzymatic labels are typically detected by providing the enzyme with a substrate and detecting the reaction product produced by the action of the enzyme on the substrate, and colorimetric labels are detected by simply visualizing the colored label.

25 The labels may be incorporated by any of a number of means well known to those of skill in the art. However, in a preferred embodiment, the label is simultaneously incorporated during the amplification step in the preparation of the sample nucleic acids. Thus, for

example, polymerase chain reaction (PCR) with labeled primers or labeled nucleotides will provide a labeled amplification product. In a preferred embodiment, transcription amplification, as described above, using a labeled nucleotide (e.g. fluorescein-labeled UTP and/or CTP) incorporates a label into the transcribed nucleic acids.

5 Alternatively, a label may be added directly to the original nucleic acid sample (e.g., mRNA, polyA mRNA, cDNA, etc.) or to the amplification product after the amplification is completed. Means of attaching labels to nucleic acids are well known to those of skill in the art and include, for example, nick translation or end-labeling (e.g. with a labeled RNA) by kinasing of the nucleic acid and subsequent attachment (ligation) of a nucleic acid linker
10 joining the sample nucleic acid to a label (e.g., a fluorophore).

In a preferred embodiment, the fluorescent modifications are by cyanine dyes e.g. Cy-3/Cy-5 dUTP, Cy-3/Cy-5 dCTP (Amersham Pharmacia) or alexa dyes (Khan, et al., 1998, *Cancer Res.* 58:5009-5013).

15 In a preferred embodiment, the two target samples used for comparison are labeled with different fluorescent dyes which produce distinguishable detection signals, for example, targets made from normal cartilage are labeled with Cy5 and targets made from mild osteoarthritis cartilage are labeled with Cy3. The differently labeled target samples are hybridized to the same microarray simultaneously. In a preferred embodiment, the labeled targets are purified using methods known in the art, e.g., by ethanol purification or column
20 purification.

25 In a preferred embodiment, the target will include one or more control molecules which hybridize to control probes on the microarray to normalize signals generated from the microarray. Preferably, labeled normalization targets are nucleic acid sequences that are perfectly complementary to control oligonucleotides that are spotted onto the microarray as described above. The signals obtained from the normalization controls after hybridization provide a control for variations in hybridization conditions, label intensity, "reading" efficiency and other factors that may cause the signal of a perfect hybridization to vary

between arrays. In a preferred embodiment, signals (e.g., fluorescence intensity) read from all other probes in the array are divided by the signal (e.g., fluorescence intensity) from the control probes, thereby normalizing the measurements.

Preferred normalization targets are selected to reflect the average length of the other

5 targets present in the sample, however, they are selected to cover a range of lengths. The normalization control(s) also can be selected to reflect the (average) base composition of the other probes in the array, however, in a preferred embodiment, only one or a few normalization probes are used and they are selected such that they hybridize well (i.e., have no secondary structure and do not self hybridize) and do not match any target molecules.

10 Normalization probes are localized at any position in the array or at multiple positions throughout the array to control for spatial variation in hybridization efficiency. In a preferred embodiment, normalization controls are located at the corners or edges of the array as well as in the middle.

Hybridization Conditions

15 Nucleic acid hybridization involves providing a denatured probe or target nucleic acid member and target nucleic acid under conditions where the probe or target nucleic acid member and its complementary target can form stable hybrid duplexes through complementary base pairing. The nucleic acids that do not form hybrid duplexes are then washed away leaving the hybridized nucleic acids to be detected, typically through detection
20 of an attached detectable label. It is generally recognized that nucleic acids are denatured by increasing the temperature or decreasing the salt concentration of the buffer containing the nucleic acids. Under low stringency conditions (e.g., low temperature and/or high salt) hybrid duplexes (e.g., DNA:DNA, RNA:RNA, or RNA:DNA) will form even where the annealed sequences are not perfectly complementary. Thus specificity of hybridization is reduced at
25 lower stringency. Conversely, at higher stringency (e.g., higher temperature or lower salt) successful hybridization requires fewer mismatches.

The invention provides for hybridization conditions comprising the Dig hybridization mix (Boehringer); or formamide-based hybridization solutions, for example as described in Ausubel et al., *supra* and Sambrook et al. *supra*.

Methods of optimizing hybridization conditions are well known to those of skill in the art (see, e.g., *Laboratory Techniques in Biochemistry and Molecular Biology*, Vol. 24: *Hybridization With Nucleic acid Probes*, P. Tijssen, ed. Elsevier, N.Y., (1993)).

Following hybridization, non-hybridized labeled or unlabeled nucleic acid is removed from the support surface, conveniently by washing, thereby generating a pattern of hybridized target nucleic acid on the substrate surface. A variety of wash solutions are known to those of skill in the art and may be used. The resultant hybridization patterns of labeled, hybridized oligonucleotides and/or nucleic acids may be visualized or detected in a variety of ways, with the particular manner of detection being chosen based on the particular label of the test nucleic acid, where representative detection means include scintillation counting, autoradiography, fluorescence measurement, calorimetric measurement, light emission measurement and the like.

Image Acquisition and Data Analysis

Following hybridization and any washing step(s) and/or subsequent treatments, as described above, the resultant hybridization pattern is detected. In detecting or visualizing the hybridization pattern, the intensity or signal value of the label will be not only be detected but quantified, by which is meant that the signal from each spot of the hybridization will be measured and compared to a unit value corresponding to the signal emitted by a known number of end labeled target nucleic acids to obtain a count or absolute value of the copy number of each end-labeled target that is hybridized to a particular spot on the array in the hybridization pattern.

Methods for analyzing the data collected from hybridization to arrays are well known in the art. For example, where detection of hybridization involves a fluorescent label, data analysis can include the steps of determining fluorescent intensity as a function of substrate

position from the data collected, removing outliers, i.e., data deviating from a predetermined statistical distribution, and calculating the relative binding affinity of the test nucleic acids from the remaining data. The resulting data is displayed as an image with the intensity in each region varying according to the binding affinity between associated oligonucleotides and/or nucleic acids and the test nucleic acids.

The following detection protocol is used for the simultaneous analysis of two cartilage samples to be compared, where each sample is labeled with a different fluorescent dye.

Each element of the microarray is scanned for the first fluorescent color. The intensity of the fluorescence at each array element is proportional to the expression level of that gene in the sample.

The scanning operation is repeated for the second fluorescent label. The ratio of the two fluorescent intensities provides a highly accurate and quantitative measurement of the relative gene expression level in the two tissue samples.

In a preferred embodiment, fluorescence intensities of immobilized target nucleic acid sequences were determined from images taken with a custom confocal microscope equipped with laser excitation sources and interference filters appropriate for the Cy3 and Cy5 fluors. Separate scans were taken for each fluor at a resolution of $225 \mu\text{m}^2$ per pixel and 65,536 gray levels. Image segmentation to identify areas of hybridization, normalization of the intensities between the two fluor images, and calculation of the normalized mean fluorescent values at each target are as described (Khan, et al., 1998, *Cancer Res.* 58:5009-5013. Chen, et al., 1997, *Biomed. Optics* 2:364-374). Normalization between the images is used to adjust for the different efficiencies in labeling and detection with the two different fluors. This is achieved by equilibrating to a value of one the signal intensity ratio of a set of internal control genes spotted on the array.

In another preferred embodiment, the array is scanned in the Cy 3 and Cy5 channels and stored as separate 16-bit TIFF images. The images are incorporated and analysed using software which includes a gridding process to capture the hybridization intensity data from

each spot on the array. The fluorescence intensity and background-subtracted hybridization intensity of each spot is collected and a ratio of measured mean intensities of Cy5 to Cy3 is calculated. A liner regression approach is used for normalization and assumes that a scatter plot of the measured Cy5 versus Cy3 intensities should have a scope of one. The average of 5 the ratios is calculated and used to rescale the data and adjust the slope to one. A post-normalization cutoff of greater than 1.0 fold up- or down-regulation is used to identify differentially expressed genes.

Following detection or visualization, the hybridization pattern is used to determine quantitative information about the genetic profile of the labeled target nucleic acid sample 10 that was contacted with the array to generate the hybridization pattern, as well as the physiological source from which the labeled target nucleic acid sample was derived. By “genetic profile” is meant information regarding the types of nucleic acids present in the sample, e.g., such as the types of genes to which they are complementary, and/or the copy number of each particular nucleic acid in the sample. From this data, one can also derive 15 information about the physiological source from which the target nucleic acid sample was derived, such as the types of genes expressed in the tissue or cell which is the physiological source of the target, as well as the levels of expression of each gene, particularly in quantitative terms.

Where one uses the subject methods to compare target nucleic acids from two or more 20 physiological sources, the hybridization patterns may be compared to identify differences between the patterns. Where arrays in which each of the different nucleic acid members corresponds to a known gene are employed, any discrepancies are related to a differential expression of a particular gene in the physiological sources being compared. Thus, the subject methods find use in differential gene expression assays, where one may use the 25 subject methods in the differential expression analysis of: (a) diseased vs. normal tissue, e.g., osteoarthritic and normal tissue, (b) tissue derived from different stages of osteoarthritis; and the like.

In a particularly preferred embodiment, where it is desired to quantify the transcription level (and thereby expression) of one or more nucleic acid sequences in a sample, the target nucleic acid sample is one in which the concentration of the mRNA transcript(s) of the gene or genes, or the concentration of the nucleic acids derived from the mRNA transcript(s), is proportional to the transcription level (and therefore expression level) of that gene. Similarly, it is preferred that the hybridization signal intensity be proportional to the amount of hybridized nucleic acid. While it is preferred that the proportionality be relatively strict (e.g., a doubling in transcription rate results in a doubling in mRNA transcript in the sample nucleic acid pool and a doubling in hybridization signal), one of skill will appreciate that the proportionality can be more relaxed and even non-linear and still provide meaningful results. Thus, for example, an assay where a 5 fold difference in concentration of the target mRNA results in a 3- to 6-fold difference in hybridization intensity is sufficient for most purposes. Where more precise quantification is required, appropriate controls are run to correct for variations introduced in sample preparation and hybridization as described herein. In addition, serial dilutions of "standard" target mRNAs are used to prepare calibration curves according to methods well known to those of skill in the art. Of course, where simple detection of the presence or absence of a transcript is desired, no elaborate control or calibration is required.

For example, if a microarray nucleic acid member is not labeled after hybridization, this indicates that the gene comprising that nucleic acid member is not expressed in either sample. If a nucleic acid member is labeled with a single color, it indicates that a labeled gene was expressed only in one sample. The labeling of a nucleic acid member comprising an array with both colors indicates that the gene was expressed in both samples. Even genes expressed once per cell are detected (1 part in 100,000 sensitivity). A difference in expression intensity in the two samples being compared is indicative of differential expression, the ratio of the intensity in the two samples being not equal to 1.0, preferably less than 0.7 or greater than 1.2, more preferably less than 0.5 or greater than 1.5.

Many human genes are expressed at different levels in cartilage of different developmental (fetal vs. mature) or disease states. In some cases, a gene is not expressed at

all in some developmental or disease states, and at high levels in others. Differential analysis of chondrocyte gene expression in differing cartilage states using an EST-based approach is used to identify genes that may play important roles in osteoarthritis pathogenesis and cartilage repair. The advantage of this method is that it can provide gene expression 5 information on a larger scale than other methods. The cDNA clones generated by this approach is useful for future functional studies of certain genes. This type of genomic-based approach can provide important novel insights into our understanding of the osteoarthritis disease process and provide for novel diagnostic, prognostic and therapeutic approaches.

Diagnostic or Prognostic Tests

10 The invention also provides for diagnostic tests for detecting osteoarthritis. The invention also provides for prognostic tests for monitoring a patient's response to therapy.

According to the method of the invention, mild, moderate, marked or severe 15 osteoarthritis is detected by obtaining a cartilage sample from a patient. In alternative embodiments, a blood or synovial fluid sample is obtained from a patient . A sample comprising nucleic acid corresponding to RNA (i.e., RNA or cDNA) is prepared from the patient cartilage (or blood or synovial fluid) sample. The sample comprising nucleic acid corresponding to RNA is hybridized to an array comprising a solid substrate and a plurality of nucleic acid members, where at least one member is differentially expressed in cartilage isolated from a patient diagnosed with mild, moderate, marked or severe osteoarthritis, as 20 compared to a "normal individual", according to the invention. According to this diagnostic test, hybridization of the sample comprising nucleic acid corresponding to RNA to one or more nucleic acid members on the array is indicative of disease.

A patient response to therapy is monitored by using a prognostic test according to the invention. In one aspect, a prognostic test according to the invention comprises obtaining a 25 cartilage sample from a patient prior to treatment, during the course of treatment and after treatment. Preferably, the patient is treated for at least 12 hours before a sample is taken. In alternative embodiments, blood or synovial fluid samples are obtained from a patient prior to

treatment, during the course of treatment and after treatment . A sample comprising nucleic acid corresponding to RNA (i.e., RNA or cDNA) is prepared from the patient cartilage (or blood or synovial fluid) samples. The samples comprising nucleic acid corresponding to RNA are hybridized to an array comprising a solid substrate and a plurality of nucleic acid 5 members, where at least one member is differentially expressed in cartilage isolated from a patient diagnosed with mild, moderate, marked or severe osteoarthritis, as compared to a normal individual, according to the invention. Arrays are selected in accordance with the diagnostic state of the patient whose treatment is being monitored. According to this prognostic test, differential hybridization of the samples comprising nucleic acid 10 corresponding to RNA isolated prior to and after treatment to one or more nucleic acid members on the array is indicative of an effective treatment. Preferably, gene expression profiles in patients being treated changes to resemble more closely gene expression profiles in patients with less severe forms of the disease or more preferably more closely resembles gene expression profiles in normal patients. The extent of change in a gene expression profile can 15 be further correlated with various therapeutic endpoints such as a decrease in the severity and/or occurrence of one or more symptoms associated with the disease.

Therapeutic Agents

A useful therapeutic agent according to the invention can increase or decrease the anabolic and/or the catabolic activity of a chondrocyte. Preferably, a therapeutic agent can 20 increase or decrease the anabolic and/or catabolic activity of a chondrocyte by greater than 1.0-fold, more preferably, 1.5-5-fold, and most preferably, 5-100-fold, as compared to an untreated chondrocyte.

In one embodiment, a therapeutic agent changes (e.g., increases or decreases) the level 25 of expression of at least one nucleic acid sequence that is differentially expressed in a chondrocyte derived from any of the following chondrocyte disease or developmental stages: fetal, normal, mild osteoarthritic, moderate osteoarthritic, marked osteoarthritic and severe osteoarthritic. Preferably, a therapeutic agent causes a change in the level of expression of a nucleic acid sequence or increase or decrease in the expression of a nucleic acid sequence that

is differentially expressed in a chondrocyte derived from any of the following chondrocyte disease or developmental stages: fetal, normal, mild osteoarthritic, moderate osteoarthritic, marked osteoarthritic and severe osteoarthritic, where the change is greater than 1.0-fold, more preferably 1.5-5-fold, and most preferably 5-100-fold, more or less than the level of 5 expression in the absence of a candidate therapeutic agent.

In another embodiment, a therapeutic agent according to the invention can ameliorate at least one of the symptoms and/or changes associated with osteoarthritis including cartilage degeneration, or pain, swelling, weakness and/or loss of functional ability in the afflicted joints, associated with cartilage degeneration.

10 The candidate therapeutic agent may be a synthetic compound, or a mixture of compounds, or may be a natural product (*e.g.* a plant extract or culture supernatant).

Candidate therapeutic agents or compounds from large libraries of synthetic or natural compounds can be screened. Numerous means are currently used for random and directed synthesis of saccharide, peptide, and nucleic acid-based compounds. Synthetic compound libraries are commercially available from a number of companies including Maybridge Chemical Co. (Trevillet, Cornwall, UK), Comgenex (Princeton, NJ), Brandon Associates (Merrimack, NH), and Microsource (New Milford, CT). A rare chemical library is available from Aldrich (Milwaukee, WI). Combinatorial libraries are available and are prepared. Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant and animal 20 extracts are available from *e.g.*, Pan Laboratories (Bothell, WA) or MycoSearch (NC), or are readily producable by methods well known in the art. Additionally, natural and synthetically produced libraries and compounds are readily modified through conventional chemical, physical, and biochemical means.

Useful compounds may be found within numerous chemical classes. Useful 25 compounds may be organic compounds, or small organic compounds. Small organic compounds have a molecular weight of more than 50 yet less than about 2,500 daltons, preferably less than about 750, more preferably less than about 350 daltons. Exemplary

classes include heterocycles, peptides, saccharides, steroids, and the like. The compounds may be modified to enhance efficacy, stability, pharmaceutical compatibility, and the like. Structural identification of an agent may be used to identify, generate, or screen additional agents. For example, where peptide agents are identified, they may be modified in a variety 5 of ways to enhance their stability, such as using an unnatural amino acid, such as a D-amino acid, particularly D-alanine, by functionalizing the amino or carboxylic terminus, e.g. for the amino group, acylation or alkylation, and for the carboxyl group, esterification or amidification, or the like.

10 A therapeutic agent, according to the invention, can be a gene corresponding to an EST sequence identified from any of the cDNA libraries constructed from cartilage of different development and disease stages.

15 Each cDNA library revealed a number of EST sequences specific to the particular stage. The ESTs are first characterized according to their putative function (Tables 2-6) and their expression is confirmed by using microarrays, as described herein. Since osteoarthritis is a chronic disease caused by an imbalance between catabolic activity and anabolic activity, that is, an increase in catabolic activity and/or a decrease in anabolic activity, normal- or fetal-specific ESTs may be important in maintaining the normal metabolic function of cartilage so as to maintain a balance between the catabolic activity and the anabolic activity. Therefore, 20 an increased expression of a full length gene sequence corresponding to one or more of these ESTs may restore the anabolic activity in disease cartilage. Therapy involving altered gene expression (e.g., gene therapy, gene disruption, antisense therapy, and the like) is useful according to the invention.

25 A full-length gene sequence corresponding to one of the normal- or fetal-specific genes is cloned by methods known in the art (e.g., Ausubel et al., John Wiley & Sons, Inc., 1997, *Current Protocols in Molecular Biology*). A cloned sequence is transfected into disease chondrocytes isolated from any stage of osteoarthritis (e.g., mild, moderate, marked, and severe). The ability of normal- or fetal-specific genes to complement the anabolic defect in the disease chondrocytes is accessed.

In one embodiment, this is achieved by examining the expression profile of disease chondrocytes transfected with a normal- or fetal-specific gene. A normal- or fetal-specific gene which is capable of restoring the expression profile of disease chondrocytes to more closely resemble that of normal or fetal chondrocytes is a useful candidate for treatment of
5 osteoarthritis.

In another embodiment, the anabolic activity of disease chondrocytes transfected with a normal- or fetal- specific gene is measured as described by Westacott et al. (1996, *Semin Arthritis Rheum*, 25:254-72). A normal- or fetal-specific gene which increases the anabolic activity is useful for treatment of osteoarthritis.

Once a therapeutic gene is defined, the gene sequence is subcloned into a vector suitable for the purpose of gene therapy. Murine leukemia virus (MLV)-based retroviral vectors are one of the most widely used gene delivery vehicles in gene therapy clinical trials and have been employed in almost 70% of approved protocols (Ali, M. et al., 1994, *Gene Ther.*, 1:367-384; Marshall, 1995, *Science*, 269:1050-1055, 1995). Other useful vectors are also known in the art (e.g., Carter and Samulski, 2000, *Int. J. Mol. Med.* 6:17-27; Lever et al., 1999, *Biochem. Soc. Trans.* 27: 841-7). Methods for gene therapy of human diseases are described in U.S. Patent Nos. 6,190,907; 6,187,305; 6,140,087; and 6,129,705, for example, the entireties of which are incorporated by reference herein.

Dosage and Administration

Therapeutic agents of the invention are administered to a patient, preferably in a biologically compatible solution or a pharmaceutically acceptable delivery vehicle, by ingestion, injection, inhalation or any number of other methods routine in the art. The dosages administered will vary from patient to patient. A "therapeutically effective dose" is determined, for example, by the level of enhancement of function (e.g., increased or decreased chondrocyte anabolic activity, or an increase or decrease in the expression of at least one nucleic acid sequence that is differentially expressed in a chondrocyte derived from
20
25

any of the following chondrocyte disease or developmental stages: fetal, normal, mild osteoarthritic, moderate osteoarthritic, marked osteoarthritic or severe osteoarthritic).

A therapeutic agent according to the invention is administered in a single dose. This dosage may be repeated daily, weekly, monthly, yearly, or as considered appropriate by the
5 treating physician.

Pharmaceutical Compositions

The invention provides for compositions comprising a therapeutic agent according to the invention admixed with a physiologically compatible carrier. As used herein, "physiologically compatible carrier" refers to a physiologically acceptable diluent such as water, phosphate buffered saline, or saline, and further may include an adjuvant. Adjuvants such as incomplete Freund's adjuvant, aluminum phosphate, aluminum hydroxide, or alum are materials well known in the art.

The invention also provides for pharmaceutical compositions. In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically acceptable carrier preparations which is used pharmaceutically.

Pharmaceutical compositions for oral administration are formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions and the like, for
20 ingestion by the patient.

Pharmaceutical preparations for oral use are obtained through a combination of active compounds with solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are carbohydrate or protein fillers such as sugars, including lactose,
25 sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethyl cellulose;

and gums including arabic and tragacanth; and proteins such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate.

Dragee cores are provided with suitable coatings such as concentrated sugar solutions,

5 which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, i.e., dosage.

Pharmaceutical preparations which are used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders such as lactose or starches, lubricants such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid paraffin, or liquid polyethylene glycol with or without stabilizers.

Pharmaceutical formulations for parenteral administration include aqueous solutions of active compounds. For injection, the pharmaceutical compositions of the invention may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hank's solution, Ringer' solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

For nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

The pharmaceutical compositions of the present invention may be manufactured in a manner known in the art, e.g. by means of conventional mixing, dissolving, granulating, dragee-making, levitating, emulsifying, encapsulating, entrapping or lyophilizing processes.

The pharmaceutical composition may be provided as a salt and are formed with many acids, including but not limited to hydrochloric, sulfuric, acetic, lactic, tartaric, malic, succinic, etc. Salts tend to be more soluble in aqueous or other protonic solvents that are the corresponding free base forms. In other cases, the preferred preparation may be a lyophilized powder in 1mM-50 mM histidine, 0.1%-2% sucrose, 2%-7% mannitol at a pH range of 4.5 to 5.5 that is combined with buffer prior to use.

After pharmaceutical compositions comprising a therapeutic agent of the invention formulated in an acceptable carrier have been prepared, they are placed in an appropriate container and labeled for treatment of an indicated condition with information including amount, frequency and method of administration.

Efficacy of Osteoarthritis Therapy Using Defined Therapeutic Agents

The efficacy of the therapy using any of the therapeutic agents according to the invention is determined by a medical practitioner. This determination may be related to alleviating osteoarthritis symptoms such as pain, swelling, weakness and loss of functional ability in the afflicted joint(s), and/or criteria for osteoarthritis diagnosis and staging described in Marshall (1996, *supra*).

The above disclosure generally describes the present invention. A more complete understanding can be obtained by reference to the following specific examples, which are provided herein for purposes of illustration only and are not intended to limit the scope of the invention.

Examples

The examples below are non-limiting and are merely representative of various aspects and features of the present invention

Example 1: RNA Extraction And Fetal cDNA Library Construction

A cDNA library was prepared from fetal cartilage. ESTs were obtained from the 5 cDNA library and evaluated to create one or more gene expression profiles for fetal chondrocytes.

Human fetal femoral cartilage RNA was extracted from pooled specimens of aborted fetuses (8-12 weeks). Samples were finely powdered under liquid nitrogen, and total RNA was extracted using TRIzol® reagent (GIBCO/BRL). Purity and integrity of RNA was assessed by absorbance at 260/280nm and agarose gel electrophoresis. The poly (A)⁺ RNA fraction was isolated by oligo-dT cellulose chromatography (Pharmacia), and 3-5 ug poly (A)⁺ RNA was used to construct a cDNA library in the λ ZAP Express vector (Stratagene). First-strand cDNA was synthesized with an Xho I-oligo (dT) adapter-primer in the presence of 5'-methyl dCTP. After second-strand synthesis and ligation of EcoRI adapters, the cDNA was digested with Xho I, resulting in cDNA flanked by EcoRI sites at the 5'-ends and Xho I sites at the 3'-ends. Digested cDNAs were size-fractionated in Sephadryl S-500 spin columns (Stratagene), then ligated into the λ ZAP Express vector predigested with EcoRI and Xho I. The resulting DNA/cDNA concatomers were packaged using Gigapack Gold packaging extracts. After titration, aliquots of primary packaging mix were stored in 7% DMSO at - 20 80°C as primary library stocks, and the rest were amplified to establish stable library stocks.

Large-scale sequencing of cDNA inserts

From the amplified λ ZAP Express library, phage plaques were plated at a density of 200-500 pfu/150 mm plate onto *Escherichia coli* XL1-blue MRF' lawn with IPTG/X-gal for color selection. Plaques were picked into 75 ul suspension media buffer (100 mM NaCl, 10 25 mM MgSO₄, 1 mM Tris, pH7.5, 0.02% gelatin). Phage elutes (5 ul) were used for PCR reactions (50 ul total volume) with 125 umol/L of each dNTP (Pharmacia), 10 pmol each of

modified T3 (5'- GCCAAGCTCGAAATTAACCCTCACTAAAG GG-3') and T7 (5'- CCAGTGAATTGTAATACGACTCACTATAGGGCG-3') primers, and 2 U of Taq DNA polymerase (Pharmacia). Reactions were cycled in a DNA Thermal Cycler (Perkin-Elmer) [denaturation at 95°C for 5 minutes, followed by 30 cycles of amplification (94°C, 45 5 seconds; 55°C, 30 seconds; 72°C, 3 minutes) and a terminal isothermal extension (72°C, 3 minutes)]. Agarose gel electrophoresis was used to assess the presence and purity of inserts. PCR products are subjected to DNA sequencing reactions using specific primers, BigDye™ Terminator Cycle Sequencing v2.0 Ready Reaction (PE Biosystems), Tris MgCl buffer and water in a thermocycler. Sequencing reactions were incubated at 94°C for 2 minutes, 10 followed by 25 cycles of 94°C, 30 seconds; 55°C, 20 seconds; and 72°C, 1 minute; and 15 cycles of 94°C, 30 seconds; and 72°C for 1 minute; and 72°C for 5 minutes. Reactions were then put on hold at 4°C until purified through methods well known in the prior art (i.e. column purification or alcohol precipitation). Automated sequencing was carried out with a PE Biosystems ABI Prism 3700 DNA Analyzer.

15 Sequences were manually edited or edited using Sequencher software (GeneCodes). All edited EST sequences were compared to the non-redundant Genbank/EMBL/DDBJ and dbEST databases using the BLAST algorithm (8). A minimum value of $P = 10^{-10}$ and nucleotide sequence identity >95% were required for assignments of putative identities for ESTs matching to known genes or to other ESTs. Construction of a non-redundant list of 20 genes represented in the EST set was done with the help of Unigene, Entrez and PubMed at the National Center for Biotechnology Information (NCBI) site (<http://www.ncbi.nlm.nih.gov/>). Relative gene expression frequency was calculated by dividing the number of EST copies for each gene by the total number of ESTs analyzed. Functional characterization of ESTs with known gene matches was made according to the 25 categories described by Hwang et al., "A Genome-Based Resource for Molecular Cardiovascular Medicine: Toward a Compendium of Cardiovascular Genes." *Circulation* 1997;96:4146-203).

A total of 13,398 ESTs were obtained from the human fetal cartilage cDNA library. Of these, 5,747 ESTs (41.8%) matched to known gene sequences, 1,855 ESTs (13.4%) matched to other ESTs, and 3,053 (22.0%) matched to mitochondrial, ribosomal, vector and cDNA/hypothetical protein sequences. The 209 ESTs (4.7%) that did not match to any known sequences were designated as novel. The remainder matched to genomic DNA sequences (1,948 ESTs, 13.8%) and repetitive sequences (586 ESTs, 4.3%).

The 13,398 EST sequences in the fetal library were characterized based on the functional classification of the 2,579 unique known genes they represented. The following table sets out the results of this analysis.

Table 2. Fetal Cartilage Library		
Putative Function	Percent (%) of Representation in the Library	Number of genes
Cell division	7.06	182
Cell signaling/communication	15.01	387
Cell structure/motility	10.90	281
Cell/organism defense	7.60	196
Gene/protein expression	22.22	573
Metabolism	14.89	384
Unclassified	22.33	576
Total known/unique genes analyzed	100.00	2,579

10

Example 2: RNA Extraction And Normal Adult cDNA Library Construction

A cDNA library was prepared from normal adult cartilage. ESTs were obtained from the cDNA library and characterized to create one or more gene expression profiles for normal adult chondrocytes.

15 Large-scale sequencing of cDNA inserts

cDNA libraries were constructed into λ TripleEx2 vector through a PCR-based method, using SMART (Switching Mechanism At 5' end of RNA Transcript) cDNA Library

Construction Kit (Clontech). Phage plaques were randomly picked and positive inserts were identified by PCR. Agarose gel electrophoresis was used to assess the presence and purity of inserts. PCR product was then subjected to automated DNA sequencing with a 5' vector-specific forward primer and sequenced by ABI PRISM 377 DNA sequencer (Perkin Elmer) and ABI PRISM 3700 DNA Analyzer (Applied Biosystems). All generated EST sequences were searched against the nonredundant Genbank/EMBL/DDBL, dbEST and GSS databases. A minimum value of $p=10^{-10}$ and nucleotide sequence identity >90% were required for assignments of putative identities for EST-matching to known genes or other ESTs. Relative EST frequency level was calculated by dividing the EST number matched to that gene into the total number of ESTs obtained from the library.

Sequences were manually edited or edited using Sequencher software (GeneCodes). All edited EST sequences were compared to the non-redundant Genbank/EMBL/DDBJ and dbEST databases using the BLAST algorithm (8). A minimum value of $P = 10^{-10}$ and nucleotide sequence identity >95% were required for assignments of putative identities for ESTs matching to known genes or to other ESTs. Construction of a non-redundant list of genes represented in the EST set was done with the help of Unigene, Entrez and PubMed at the National Center for Biotechnology Information (NCBI) site (<http://www.ncbi.nlm.nih.gov/>). Relative gene expression frequency was calculated by dividing the number of EST copies for each gene by the total number of ESTs analyzed. Functional characterization of ESTs with known gene matches was made according to the categories described by Hwang et al (Hwang DM, Dempsey AA, Wang RX, Rezvani M, Barrans JD, Dai KS, et al. A Genome-Based Resource for Molecular Cardiovascular Medicine: Toward a Compendium of Cardiovascular Genes. *Circulation* 1997;96:4146-203).

A total of 17,151 ESTs were obtained from the normal cartilage cDNA library. Of these, 6,755 ESTs (44.2%) matched to 2,518 known genes. 1.4% (132 ESTs) showed no significant match and were thus designated as novel. Characterization of the 17,151 EST sequences based on functional classification of known/unique genes resulted in the following table:

Table 3. Normal Adult Cartilage Library

Putative function	Percent (%) of representation in the library	Number of genes
Cell division	6.13	160
Cell signaling/communication	13.52	353
Cell structure/motility	9.00	235
Cell/organism defense	7.51	196
Gene/protein expression	20.08	524
Metabolism	13.14	343
Unclassified	27.09	707
Total known/unique genes analyzed	100.00	2518

Example 3: RNA Extraction and cDNA Library Construction From Mild Osteoarthritic Chondrocytes and Severe Osteoarthritic Chondrocytes

A cDNA library was prepared from mild osteoarthritic cartilage and severe osteoarthritic cartilage. ESTs were obtained from the cDNA libraries and characterized to create one or more gene expression profiles for mild osteoarthritic chondrocytes and severe osteoarthritic chondrocytes.

Articular cartilage was obtained during either arthroscopic knee surgery or total knee replacement. The cartilage samples were obtained from either areas of very early cartilage degeneration (mild) or from sites of end stage disease (severe). cDNA libraries were constructed as described for normal adult samples (Example 2).

Large-scale sequencing of cDNA inserts

cDNA libraries were constructed into λ TripleEx2 vector through a PCR-based method, using SMART (Switching Mechanism At 5' end of RNA Transcript) cDNA Library Construction Kit (Clontech). Phage plaques were randomly picked and positive inserts were identified by PCR. Agarose gel electrophoresis was used to assess the presence and purity of inserts. PCR product was then subjected to automated DNA sequencing with a 5' vector-

specific forward primer and sequenced by ABI PRISM 377 DNA sequencer (Perkin Elmer) and ABI PRISM 3700 DNA Analyzer (Applied Biosystems). All generated EST sequences were searched against the nonredundant Genbank/EMBL/DDBL, dbEST and GSS databases. A minimum value of $p=10^{-10}$ and nucleotide sequence identity >90% were required for 5 assignments of putative identities for EST-matching to known genes or other ESTs. Relative EST frequency level was calculated by dividing the EST number matched to that gene into the total number of ESTs obtained from the library.

Sequences were manually edited or edited using Sequencher software (GeneCodes). All edited EST sequences were compared to the non-redundant Genbank/EMBL/DDBJ and 10 dbEST databases using the BLAST algorithm (8). A minimum value of $P = 10^{-10}$ and nucleotide sequence identity >95% were required for assignments of putative identities for ESTs matching to known genes or to other ESTs.

Construction of a non-redundant list of genes represented in the EST set was done with the help of Unigene, Entrez and PubMed at the National Center for Biotechnology 15 Information (NCBI) site (<http://www.ncbi.nlm.nih.gov/>). Relative gene expression frequency was calculated by dividing the number of EST copies for each gene by the total number of ESTs analyzed.

Functional characterization of ESTs with known gene matches was made according to the categories described by Hwang et al (Hwang DM, Dempsey AA, Wang RX, Rezvani M, 20 Barrans JD, Dai KS, et al. A Genome-Based Resource for Molecular Cardiovascular Medicine: Toward a Compendium of Cardiovascular Genes. *Circulation* 1997;96:4146-203).

A total of 12,651 ESTs and 14,222 ESTs were obtained from the mild and severe OA 25 cDNA libraries (Table 5 and Table 6), respectively. About 43% of mild and 51% of severe OA ESTs matched to known genes in the database. Approximately 2.6% and 1.5% of the ESTs, respectively, resulted in no significant match and were thus designated as novel.

Characterization of the 12,651 EST sequences from the mild OA library and of the 14,222 EST sequences from the severe OA library based on functional characterization of the known genes represented resulted in the following tables:

Table 4. Mild OA Cartilage Library		
Putative Function	Percent (%) of Representation in the Library	Number of Genes
Cell division	6.39	127
Cell signaling/communication	15.31	304
Cell structure/motility	9.16	182
Cell/organism defense	8.41	167
Gene/protein expression	21.60	429
Metabolism	13.95	277
Unclassified	22.76	452
Total known/unique genes analyzed	100.00	1938

Table 5. Severe OA Cartilage Library		
Putative Function	Percent (%) of Representation in the Library	Number of genes
Cell division	6.81	157
Cell signaling/communication	14.14	326
Cell structure/motility	8.50	196
Cell/organism defense	7.98	184
Gene/protein expression	22.94	529
Metabolism	13.53	312
Unclassified	23.94	552
Total known/unique genes analyzed	100.00	2256

Example 4: Identification Of Differentially Expressed Genes In Fetal, Normal Mild Osteoarthritic, And Severe Osteoarthritic Cartilage

Genes that are differentially expressed as defined herein between normal, mild, severe and fetal cartilage were identified through relative EST frequency analysis (see Figure 6). Of 5 the 5,807 known unique genes identified in Figure 6, 405 genes were found to be expressed in all four tissue types. Examples of possible subanalysis are shown in Figures 15 and 16. Some of these genes with particularly marked differential expression are shown in Figure 4. The relative frequency of ESTs representing collagens (Figures 2 and 3) and selected extracellular matrix proteins (see Figure 1) were also analyzed.

10 Example 5: Microarray Construction

A microarray according to the invention was constructed as follows.

PCR products (~40 ul) of cDNA clones from OA cartilage cDNA libraries, in the same 96-well tubes used for amplification, are precipitated with 4 ul (1/10 volume) of 3M sodium acetate (pH 5.2) and 100 ul (2.5 volumes) of ethanol and stored overnight at -20°C. They are then centrifuged at 3,300 rpm at 4°C for 1 hour. The obtained pellets were washed with 50 ul ice-cold 70% ethanol and centrifuged again for 30 minutes. The pellets are then air-dried and resuspended well in 50% dimethylsulfoxide (DMSO) or 20ul 3X SSC overnight. The samples are then deposited either singly or in duplicate onto Gamma Amino Propyl Silane (Corning CMT-GAPS or CMT-GAP2, Catalog No. 40003, 40004) or 20 polylysine-coated slides (Sigma Cat. No. P0425) using a robotic GMS 417 or 427 arrayer (Affymetrix, CA). The boundaries of the DNA spots on the microarray are marked with a diamond scriber. The invention provides for arrays where 10-20,000 PCR products are spotted onto a solid support to prepare an array.

The arrays are rehydrated by suspending the slides over a dish of warm particle free 25 ddH₂O for approximately one minute (the spots will swell slightly but not run into each other) and snap-dried on a 70-80°C inverted heating block for 3 seconds. DNA is then UV

crosslinked to the slide (Stratagene, Stratalinker, 65 mJ – set display to “650” which is 650 x 100 uJ) or baked at 80C for two to four hours. The arrays are placed in a slide rack. An empty slide chamber is prepared and filled with the following solution: 3.0 grams of succinic anhydride (Aldrich) is dissolved in 189 ml of 1-methyl-2-pyrrolidinone (rapid addition of reagent is crucial); immediately after the last flake of succinic anhydride dissolved, 21.0 ml of 0.2 M sodium borate is mixed in and the solution is poured into the slide chamber. The slide rack is plunged rapidly and evenly in the slide chamber and vigorously shaken up and down for a few seconds, making sure the slides never leave the solution, and then mixed on an orbital shaker for 15-20 minutes. The slide rack is then gently plunged in 95°C ddH₂O for 2 minutes, followed by plunging five times in 95% ethanol. The slides are then air dried by allowing excess ethanol to drip onto paper towels. The arrays are then stored in the slide box at room temperature until use.

Example 6: Target Nucleic acid Preparation and Hybridization

Preparation of Fluorescent DNA Probe from mRNA

Fluorescently labeled target nucleic acid samples are prepared for analysis with an array of the invention.

2 µg Oligo-dT primers are annealed to 2 ug of mRNA isolated from a cartilage sample from patient diagnosed with osteoarthritis or suspected of having osteoarthritis in a total volume of 15 ul, by heating to 70°C for 10 min, and cooled on ice. The mRNA is reverse transcribed by incubating the sample at 42°C for 1.5-2 hours in a 100 µl volume containing a final concentration of 50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM MgCl₂, 25 mM DTT, 25 mM unlabeled dNTPs, 400 units of Superscript II (200 U/uL, Gibco BRL), and 15 mM of Cy3 or Cy5 (Amersham). RNA is then degraded by addition of 15µl of 0.1N NaOH, and incubation at 70°C for 10 min. The reaction mixture is neutralized by addition of 15µl of 0.1N HCL, and the volume is brought to 500µl with TE (10mM Tris, 1mM EDTA), and 20 µg of Cot1 human DNA (Gibco-BRL) is added.

The labeled target nucleic acid sample is purified by centrifugation in a Centricon-30 micro-concentrator (Amicon). If two different target nucleic acid samples (e.g., two samples derived from different patients) are being analyzed and compared by hybridization to the same array, each target nucleic acid sample is labeled with a different fluorescent label (e.g., 5 Cy3 and Cy5) and separately concentrated. The separately concentrated target nucleic acid samples (Cy3 and Cy5 labeled) are combined into a fresh centricon, washed with 500 μ l TE, and concentrated again to a volume of less than 7 μ l. 1 μ L of 10 μ g/ μ l polyA RNA (Sigma, #P9403) and 1 μ l of 10 μ g/ μ l tRNA (Gibco-BRL, #15401-011) is added and the volume is adjusted to 9.5 μ l with distilled water. For final target nucleic acid preparation 2.1 μ l 20XSSC 10 (1.5M NaCl, 150mM NaCitrate (pH8.0)) and 0.35 μ l 10%SDS is added.

Hybridization

Labeled nucleic acid is denatured by heating for 2 min at 100°C, and incubated at 37°C for 20-30 min before being placed on a nucleic acid array under a 22mm x 22mm glass cover slip. Hybridization is carried out at 65°C for 14 to 18 hours in a custom slide chamber 15 with humidity maintained by a small reservoir of 3XSSC. The array is washed by submersion and agitation for 2-5 min in 2X SSC with 0.1%SDS, followed by 1X SSC, and 0.1X SSC. Finally, the array is dried by centrifugation for 2 min in a slide rack in a Beckman GS-6 tabletop centrifuge in Microplus carriers at 650 RPM for 2 min.

Example 7: Signal Detection And Data Generation

20 Following hybridization of an array with one or more labeled target nucleic acid samples, arrays are scanned immediately using a GMS Scanner 418 and Scanalyzer software (Michael Eisen, Stanford University), followed by GeneSpring software (Silicon Genetics, CA) analysis. Alternatively, a GMS Scanner 428 and Jaguar software may be used followed by GeneSpring software analysis.

25 If one target nucleic acid sample is analyzed, the sample is labeled with one fluorescent dye (e.g., Cy3 or Cy5).

After hybridization to a microarray as described in Example 6, fluorescence intensities at the associated nucleic acid members on the microarray are determined from images taken with a custom confocal microscope equipped with laser excitation sources and interference filters appropriate for the Cy3 or Cy5 fluors.

5 The presence of Cy3 or Cy5 fluorescent dye on the microarray indicates hybridization of a target nucleic acid and a specific nucleic acid member on the microarray. The intensity of Cy3 or Cy5 fluorescence represents the amount of target nucleic acid which is hybridized to the nucleic acid member on the microarray, and is indicative of the expression level of the specific nucleic acid member sequence in the target sample.

10 When two target nucleic acid samples are being analyzed and compared (e.g., mild osteoarthritic vs severe osteoarthritic), one target nucleic acid sample (for example, mild osteoarthritic) is labeled with fluorescent dye Cy3, the other target nucleic acid sample (for example, severe osteoarthritis) is labeled with fluorescent dye Cy5.

15 After hybridization as described in Example 6, fluorescence intensities at the associated nucleic acid members on the microarray are determined from images taken with a custom confocal microscope equipped with laser excitation sources and interference filters appropriate for the Cy3 and Cy5 fluors. Separate scans are taken for each fluor at a resolution of $225 \mu\text{m}^2$ per pixel and 65,536 gray levels. Normalization between the images is used to adjust for the different efficiencies in labeling and detection with the two different fluors.

20 This is achieved by manual matching of the detection sensitivities to bring a set of internal control genes to nearly equal intensity followed by computational calculation of the residual scalar required for optimal intensity matching for this set of genes.

The presence of Cy3 or Cy5 fluorescent dye on the microarray indicates hybridization of a target nucleic acid and a specific nucleic acid member on the microarray. The intensities 25 of Cy3 or Cy5 fluorescence represent the amount of target nucleic acid which is hybridized to the nucleic acid member on the microarray, and is indicative of the expression level of the specific nucleic acid member sequence in the target sample. If a nucleic acid member on the

array shows no color, it indicates that the gene in that element is not expressed in either sample. If a nucleic acid member on the array shows a single color, it indicates that a labeled gene is expressed only in that cell sample. The appearance of both colors indicates that the gene is expressed in both tissue samples. The ratios of Cy3 and Cy5 fluorescence intensities, 5 after normalization, are indicative of differences of expression levels of the associated nucleic acid member sequence in the two samples for comparison. A ratio of expression intensity not equal to 1.0 is used as an indication of differential gene expression.

10 The array is scanned in the Cy 3 and Cy5 channels and stored as separate 16-bit TIFF images. The images are incorporated and analysed using Scalyzer software which includes a gridding process to capture the hybridization intensity data from each spot on the array. The fluorescence intensity and background-subtracted hybridization intensity of each spot is collected and a ratio of measured mean intensities of Cy5 to Cy3 is calculated. A liner regression approach is used for normalization and assumes that a scatter plot of the measured Cy5 versus Cy3 intensities should have a scope of one. The average of the ratios is calculated 15 and used to rescale the data and adjust the slope to one. A post-normalization cutoff of a ratio not equal to 1.0 is used to identify differentially expressed genes.

20 Analysis of a microarray comprising some of the sequences in Figure 14, resulted in 36 candidate upregulated genes in the mild OA library that showed a greater than 2-fold median ratio and 47 candidate downregulated genes that showed a less than 0.2-fold median ratio (Figures 9 and 10, respectively,). A total of 38 candidate upregulated genes were also identified in the severe OA library that showed a greater than 2-fold median ratio and 51 candidate downregulated genes that showed a less than 0.2-fold median ratio (Figures 11 and 12, respectively,). According to this embodiment, the microarray was hybridized with a 25 target nucleic acid sample derived from an individual diagnosed with mild osteoarthritis and a target nucleic acid sample derived from an individual diagnosed with severe osteoarthritis. As would be clear to a person skilled in the art, similar analysis can be performed for any of the sequences identified in Figure 13, or the sequences identified in Figure 6A which correspond to the genes disclosed in Figure 6 using the methods disclosed herein.

Example 8: Chondrocyte-Specific Gene Microarray And Diagnosis Microarray Construction

A collection of nucleic acid members are spotted on a glass slide as described in Example 5 for the construction of a mild OA diagnosis microarray. A collection of nucleic acid members are spotted on a glass slide as described in Example 5 for the construction of a severe OA diagnosis microarray. A collection of chondrocyte specific nucleic acid members are spotted on a glass slide as described in Example 5 for the construction of a chondrocyte-specific gene microarray. The nucleic acid members spotted onto the microarrays described are selected from those named in Figures 6B, 6C, 6D and 6E.

10 Example 9: Diagnosis

Target nucleic acid samples are prepared from cartilage RNA extracts of an individual (as described in Example 6) and hybridized to a microarray comprising a collection of nucleic acid members where at least one member is differentially expressed in cartilage isolated from a patient diagnosed with mild, moderate, marked or severe osteoarthritis, as compared to cartilage isolated from a normal individual as defined herein (as described in Example 6). A hybridization pattern is generated and analyzed as in Example 7. For example, the hybridization of target nucleic acid samples to one or more nucleic acid members on the microarray comprising a collection of nucleic acid members where at least one member is differentially expressed in mild osteoarthritis cartilage as compared to a normal individual is indicative of a mild osteoarthritis of the individual from whom the target nucleic acid sample is derived. The hybridization of target nucleic acid samples to one or more nucleic acid members on the microarray comprising a collection of nucleic acid members differentially expressed in severe osteoarthritis cartilage as compared to the normal individual is indicative of severe osteoarthritis of the individual from whom the target nucleic acid sample is derived.

25 Example 10: Therapeutic Agent Screening

A candidate therapeutic agent that increases or decreases the expression of one or more nucleic acid sequences that are differentially expressed in a chondrocyte derived from

any of the following chondrocyte disease or developmental stages: fetal, mild osteoarthritic, moderate osteoarthritic, marked osteoarthritic and severe osteoarthritic, is screened according to the following method.

Chondrocytes are isolated from a “normal” individual and are incubated in the presence and absence of a candidate agent for varying amounts of time (i.e., 30 min, 1 hr, 5 hr, 24 hr, 48 hr and 96 hrs). When screening for therapeutic genes, a clone of a full gene sequence corresponding to an EST in Figure 6A or Figure 13 is used to transfect chondrocytes. The transfected chondrocytes are cultured for varying amounts of time (i.e., 1, 2, 3, 5, 7, 10, or 14 days). Following incubation, target nucleic acid samples are prepared from the chondrocytes and hybridized to a nucleic acid probe corresponding to a nucleic acid sequence which is differentially expressed in a chondrocyte derived from at least any two of the following of: fetal, normal, mild osteoarthritic, moderate osteoarthritic and severe osteoarthritic. The nucleic acid probe is labeled, for example with a radioactive label, according to methods well-known in the art and described herein. Hybridization is carried out by northern blot, for example as described in Ausubel et al., *supra* or Sambrook et al., *supra*). The differential hybridization, as defined herein, of the probe to the target nucleic acid samples from normal relative to RNA from any one of fetal, mild osteoarthritic, moderate osteoarthritic, marked osteoarthritic and severe osteoarthritic is indicative of the level of expression of RNA corresponding to a differentially expressed chondrocyte specific nucleic acid sequence. A change in the level of expression of the probe sequence as a result of the incubation step in the presence of the candidate agent, is indicative of an agent that increases or decreases the expression of the corresponding chondrocyte specific nucleic acid sequence.

Example 11: Assessing The Integrity Of Cartilage RNA Isolated Post-Mortem

The following Baboon cartilage study was preformed to evaluate the quality of freshly isolated RNA and RNA isolated at various times post-mortem.

Nine vials of baboon cartilage were obtained, and stored in liquid nitrogen till use.

Baboon cartilage from each vial was weighed and finely powdered under liquid nitrogen. The sample was then homogenized in TRIzol® reagent (0.1g/ml TRIzol®) and total RNA was extracted. The quantity of RNA was calculated according to the OD₂₆₀ value. The appearance of two sharp bands on the RNA gel indicated that the RNA was of good 5 quality.

RT-PCR was performed for the gene expression of collagen type II (COL2A1), B-actin and GAPDH, using 0.1ug total RNA from each sample.

The RNA gel pattern clearly shows that the RNA was not degraded up to 12 hours post mortem (Table 7). Therefore stable RNA should be expected from the biopsy sample 10 within 12 hours after death.

Table 7. Integrity Of Cartilage RNA Isolated Post-Mortem

Sample No.	Time Taken	Weight (g)	Total RNA (ug) - Based on OD260	RNA Gel (non Dil)	Col2A1	β -actin	GAPDH
1	Fresh	0.175	8	OK	++	++	++
2	1hr pm	0.29	9	OK	++	++	++
3	2hr	0.29	11.36	OK	++	+/-	+/-
4	3hr	0.25	2.8	OK	++	+/-	+/-
5	6hr	0.53	8.0	OK	++	+	+/-
6	8hr	0.18	5.26	OK	++	+	-
7	10hr	0.38	9.35	OK	++	+	+/-
8	12hr	0.20	6.7	OK	++	+/-	-
9	24hr	0.41	9.35	SMEAR	+/-	-	-

Collagen type II is abundant and specific to normal articular cartilage. Its mRNA level was comparable among all the samples except #9 (24 hours post-mortem). It should be noted that samples taken earlier will better reflect the natural *in vivo* state.

**Example 12. Expressed Sequence Tags (ESTs) Analysis of Human Chondrocyte Gene
5 Expression in Mild and Severe Osteoarthritic Cartilage**

Large-scale partial sequencing of cDNA libraries obtained from human fetal cartilage was performed to identify expressed sequence tags (ESTs) corresponding to genes that might play critical roles in OA progression. Large scale sequencing of cDNA libraries from human normal, mild and severe OA cartilage was also performed and a total of over 44,000 ESTs
10 from the three cDNA libraries were analyzed.

Normal cartilage was obtained from the donor program of Department of Orthopaedics and Rehabilitation, University of Miami. OA cartilage samples were obtained from either areas of very early cartilage degeneration (mild) or from sites of end stage disease (severe) during either arthroscopic knee surgery or total knee replacement. Total RNA from cartilage was extracted using TRIzol® reagent (GIBCO). cDNA libraries were constructed into λ TriplEx2 vector through a PCR-based method, using SMART (Switching Mechanism At 5' end of RNA Transcript) cDNA Library Construction Kit (Clontech) as described above. Phage plaques were randomly picked and positive inserts were identified by PCR. Agarose gel electrophoresis was used to assess the presence and purity of inserts. PCR product was
20 then subjected to automated DNA sequencing with a 5'vector-specific forward primer and sequenced by ABI PRISM 377 DNA sequencer (Perkin Elmer) and ABI PRISM 3700 DNA Analyzer (Applied Biosystems). All generated EST sequences were searched against the nonredundant Genebank/EMBL/DDBL, dbEST and GSS databases. A minimum value of $p=10^{-10}$ and nucleotide sequence identity >90% were required for assignments of putative
25 identities for EST-matching to known genes or other ESTs. Relative EST frequency level was calculated by dividing the EST number matched to that gene into the total number of ESTs obtained from the library.

A total of 17,151 ESTs, 12,651 ESTs and 14,222 ESTs were obtained from normal, mild and severe OA cDNA libraries respectively and used for gene expression profiling. About 44% of the total ESTs from these three cDNA libraries matched to known genes in the database, and about 0.9% of the ESTs (409) resulted in no significant match to known sequences and were thus designated as novel. Nonredundancy analysis of the known gene matches resulted in the identification of 2,518 unique genes in normal, 1,938 in mild and 2,256 in severe OA cartilage. Differentially expressed known genes amongst fetal (22), normal, mild, and severe OA cartilage (23) were identified by examining relative EST frequency levels as shown in Figure 6.

Some of the genes with particularly marked differential expression are shown in Figure 4 provided herewith. Heat shock protein 90 (HSP90) was the gene with the most abundant ESTs matches in mild OA. Its transcript level was low in fetal cartilage. Beta-2 microglobulin (B2M) level was higher in diseased cartilage than normal cartilage, and significantly higher in diseased cartilage than in fetal cartilage. Its EST levels in mild and severe OA were similar. Osteoblast specific factor 2 (OSF-2p1) was highly expressed in severe OA compared to fetal, mild and normal cartilage. Another differentially expressed gene was megakaryocyte stimulating factor (MSF, also known as superficial zone protein, or proteoglycan 4). It had a significantly higher expression in mild OA than in severe OA.

The relative frequency of ESTs representing the collagens was also analyzed as shown in Figure 3.

Noncollagenous matrix protein profiles showed higher EST levels of decorin (DCN), fibronectin (FN), lumican (LUM) and matrix G1a protein (MGP) in both mild and severe OA cartilage as shown in Figures 1 and 4 provided herewith.

Example 13. Microarray Analysis Of Beta-2 Microglobulin (B2M) Expression In Human Osteoarthritis

As discussed above, Beta-2 microglobulin (B2M) had a high EST expression level in mild and severe OA cartilage. B2M is a nonglycosylated polypeptide that is elevated in

inflammatory and malignant diseases. It has been shown to induce stromelysin and cyclooxygenase-2 synthesis in human synovial fibroblasts (24, 25).

B2M expression during different stages of osteoarthritis was evaluated. Human OA synovial fluid (SF) was collected from human knee joint by aspiration at arthroscopy or total 5 knee replacement. Normal samples were collected from volunteers with no history of knee injury or arthritis. Organ culture was performed as follows: human severe OA cartilage slices were cultured at one slice/well in a 24-well plate in DMEM (Dulbecco's modified Eagle medium), with 10% FCS, 100 units/ml penicillin and 100 mg/ml streptomycin (DMEM++) at 37°C in a humidified atmosphere of 5% CO₂. Cultured medium (20 ul) was then collected at 10 different time points for B2M testing. B2M levels in synovial fluid and cartilage organ cultured medium were measured using a B2M enzyme immunoassay test kit (ALPCO). Statistical significance was assessed by Student's t-test with P values less than 0.05 being considered significant. Cell culture of chondrocytes from patients with severe OA was performed as follows. Chondrocytes were derived from cartilage from patients with severe 15 OA through collagenase type II digestion. Cells were then seeded at 6.5 X 10⁴/well (3.2 X 10⁴/ml) in a 6-well plate and treated with or without 10 ug/ml B2M (Sigma) for 72 hr. Microarrays containing 5184 chondrocyte-specific cDNA clones were used for gene expression profiling.

The average B2M levels detected in normal (nor), mild (mioa), moderate (mooa), 20 marked (maoa) and severe OA (seoa) synovial fluid are shown in Figure 17. B2M in osteoarthritis synovial fluid is significantly higher than that in normal. However, no significant difference was found in B2M levels among different osteoarthritis stages.

To assess if chondrocytes contribute B2M secretion, medium from cultured severe OA cartilage was collected and tested for B2M. Figure 18 shows the release of B2M is detectable 25 after 24 hour culture and continues to increase during the 72 hour study period. At 72 hours, the accumulation of B2M was about 2.1 ug/g cartilage. Similar results were obtained across three experimental runs, each using cartilage from a different donor.

Genes regulated by B2M were detected through microarray technology as described above. Figure 19 shows a black and white representation of a two-color fluorescent scan. Cy3 labeling (which would appear as green spots) correspond to genes preferentially expressed in non-B2M treated chondrocytes, while Cy5 labeling (which would appear as reddish spots) represent genes preferentially expressed in B2M treated chondrocytes. Genes expressed at approximately equal levels would appear as yellow spots. The identity of genes was determined by the location of nucleic acid members on the array. Some of the genes that were up or down-regulated at least two-fold by B2M are listed in Table 8.

Table 8. Genes Regulated by B2M

Up-Regulated	Down-Regulated
Adrenomedullin	hypothetical protein (KIAA0102)
chitinase precursor=YKL-39	intersectin short form
collagen type III, alpha 1	KARP-1 binding protein 2 (KAB2)
manganese superoxide dismutase (SOD-2)	peripheral myelin protein 22 (PMP22)
syntaxin 7	putative GTP binding protein

Variations, modifications, and other implementations of what is described herein will occur to those of ordinary skill in the art without departing from the spirit and scope of the invention. The references provided below are incorporated herein by reference in their entireties.

References

1. Zaleske DJ. Cartilage and Bone Development. Instr Course Lect 1998;47:461-
2. Buckwalter JA, Mankin HJ. Articular Cartilage: Tissue Design and Chondrocyte-Matrix Interactions. Instr Course Lect 1998;47:477-86.
3. Westacott CI, Sharif M. Cytokines in Osteoarthritis: Mediators or Markers of Joint Destruction? Semin Arthritis Rheum 1996;25:254-72

4. Adams MD, Kerlavage AR, Fleischmann RD, Fuldner RA, Bult CJ, Lee NH, et al. Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. *Nature* 1995;377 Suppl:3-174.

5. Hwang DM, Dempsey AA, Wang RX, Rezvani M, Barrans JD, Dai KS, et al. A Genome-Based Resource for Molecular Cardiovascular Medicine: Toward a Compendium of Cardiovascular Genes. *Circulation* 1997;96:4146-203.

6. Mao M, Fu G, Wu JS, Zhang QH, Zhou J, Kan LX, et al. Identification of genes expressed in human CD34⁺ hematopoietic stem/progenitor cells by expressed sequence tags and efficient full-length cDNA cloning. *Proc Natl Acad Sci* 1998;95:8175-80.

7. Hillier LD, Lennon G, Becker M, Bonaldo MF, Chiapelli B, Chissoe S, et al. Generation and analysis of 280,000 human expressed sequence tags. *Genome Res.* 1996;6:807-28.

8. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215:403-10.

9. Mundlos S, Zabel B. Developmental Expression of Human Cartilage Matrix Protein. *Dev Dyn* 1994;199:241-52.

10. Nakamura S, Kamihagi K, Satakeda H, Katayama M, Pan H, Okamoto H, et al. Enhancement of SPARC (osteonectin) synthesis in arthritic cartilage. Increased levels in synovial fluids from patients with rheumatoid arthritis and regulation by growth factors and cytokines in chondrocyte cultures. *Arthritis Rheum* 1996;39:539-51.

11. Eyre DR, The Collagens of Articular Cartilage. *Semin Arthritis Rheum* 1991;21 (3 Suppl 2):2-11.

12. Okihana H, Yamada K. Preparation of a cDNA Library and Preliminary Assessment of 1400 Genes from Mouse Growth Cartilage. *J Bone Miner Res* 1999;14:304-10.

13. Morrison EH, Ferguson MWJ, Bayliss MT, Archer CW. The developmental of articular cartilage: I. The spatial and temporal patterns of collagen types. *J Anat* 1996;189:9-22.

14. Treilleux I, Mallein-Gerin F, le Guellec D, Herbage D. Localization of the Expression of Type I, II, III Collagens, and Aggrecan Core Protein Genes in Developing Human Articular Cartilage. *Matrix* 1992;12:221-32.

5 15. Eyre DR, Wu JJ, Niyibizi C. The collagens of bone and cartilage: Molecular diversity and supramolecular assembly. In Cohn DV, Glorieux FH, Martin TJ, editors. *Calcium Regulation and Bone Metabolism*. Amsterdam. The Netherlands: Elsevier; 1990. p. 188-94.

10 16. Birnbacher R, Amann G, Breitschopf H, Lassmann H, Suchanek G, Heinz-Erian P. Cellular localization of insulin-like growth factor II mRNA in the human fetus and the placenta: detection with a digoxigenin-labeled cRNA probe and immunocytochemistry. *Pediatr Res* 1998;43:614-20.

15 17. Wang E, Wang J, Chin E, Zhou J, Bondy CA. Cellular patterns of insulin-like growth factor system gene expression in murine chondrogenesis and osteogenesis. *Endocrinology* 1995; 136:2741-51.

20 18. van Kleffens M, Groffen C, Rosato RR, van den Eijnde SM, van Neck JW, Lindenbergh-Kortleve DJ, et al. mRNA expression patterns of the IGF system during mouse limb bud development, determined by whole mount *in situ* hybridization. *Mol Cell Endocrinol* 1998;138:151-61.

19. Braulke T, Gotz W, Claussen M. Immunohistochemical localization of insulin-like growth factor binding protein-1, -3, and -4 in human fetal tissues and their analysis in media from fetal tissue explants. *Growth Regul* 1996;6:55-65.

20. Kessler E, Takahara K, Biniaminov L, Brusel M, Greenspan DS. Bone Morphogenetic Protein-1: The Type I Procollagen C-Proteinase. *Science* 1996;271:360-2.

21. Ausubel et al., John Weley & Sons, Inc., 1997, Current Protocols in Molecular Biology

5 22. Marshall, K. et al., 2000, 46th Annual Meeting, ORS, paper No. 919.

23. Kumar, S., et al., 2000, 46th Annual Meeting, ORS, paper No. 1031.

24. Marshall K., et al., 2002, 48th Annual meeting, ORS (submitted).

25. Migita K., et al., *Biochem Biophys Res Commun* 1997, 239:621-625.

26. Migita K., et al., *Kidney Int* 1999, 55:572-578.

10
What is claimed is: